(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 319 718 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
18.06.2003 Bulletin 2003/25

(51) Int Cl.[7]: **C12Q 1/68**

(21) Application number: **01204912.8**

(22) Date of filing: **14.12.2001**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **Keygene N.V.**
6700 AE Wageningen (NL)

(72) Inventor: **Van Eijk, Michael J. T.**
3993 ST Houten (NL)

(74) Representative: **van Westenbrugge, André et al**
Nederlandsch Octrooibureau
P.O. Box 29720
2502 LS The Hague (NL)

(54) High throughput analysis and detection of multiple target sequences

(57) Method for the high throughput separation and detection of a multiplicity of target sequences in a multiplicity of samples comprising subjecting each sample to a ligation-dependent amplification assay followed by a multiple injection step comprising the consecutive and/or simultaneous injection of a multiplicity of samples, for instance in a multichannel electrophoretic device.

EP 1 319 718 A1

Description

Field of the invention

[0001] The present invention relates to the field of biotechnology. In particular the present invention provides a method for the high throughput separation and detection of nucleotide sequences, and the use of the method in the discrimination and identification of target sequence such as single nucleotide polymorphisms. The invention further provides for probes that are capable of hybridising to the target sequence of interest, primers for the amplification of ligated probes, use of these probes and primers in the identification and/or detection of nucleotide sequences that are related to a wide variety of genetic traits and genes and kits of primers and/or probes suitable for use in the method according to the invention.

Background of the invention

[0002] There is a rapidly growing interest in the detection of specific nucleic acid sequences. This interest has not only arisen from the recently disclosed draft nucleotide sequence of the human genome and the presence therein, as well as in the genomes of many other organisms, of an abundant amount of single nucleotide polymorphisms (SNP), but also from marker technologies such as AFLP. The recognition that the presence of single nucleotide substitutions (and other types of genetic polymorphisms such as small insertion/deletions; indels) in genes provide a wide variety of information has also attributed to this increased interest. It is now generally recognised that these single nucleotide substitutions are one of the main causes of a significant number of monogenically and multigenically inherited diseases, for instance in humans, or are otherwise involved in the development of complex phenotypes such as performance traits in plants and livestock species. Thus, single nucleotide substitutions are in many cases also related to or at least indicative of important traits in humans, plants and animal species.

[0003] Analysis of these single nucleotide substitutions and indels will result in a wealth of valuable information, which will have widespread implications on medicine and agriculture in the widest possible terms. It is for instance generally envisaged that these developments will result in patient-specific medication. To analyse these genetic polymorphisms, there is a growing need for adequate, reliable and fast methods that enable the handling of large numbers of samples and large numbers of (predominantly) SNPs in a high throughput fashion, without significantly compromising the quality of the data obtained.

[0004] Even though a wide diversity of high-throughput detection platforms for SNPs exist at present (such as fluorometers, DNA microarrays, mass-spectrometers and capillary electrophoresis instruments), the major limitation to achieve cost-effective high throughput detection is that a robust and efficient multiplex amplification technique for non-random selection of SNPs is currently lacking to utilise these platforms efficiently, which results in suboptimal use of these powerful detection platforms and/or high costs per datapoint.

[0005] Specifically, using common amplification techniques such as the PCR technique it is possible to amplify a limited number of target sequences by combining the corresponding primer pairs in a single amplification reaction but the number of target sequences that can be amplified simultaneously is small and extensive optimisation may be required to achieved similar amplification efficiencies of the individual target sequences. One of the solutions to multiplex amplification is to use a single primer pair for the amplification of all target sequences, which requires that all targets must contain the corresponding primer-binding sites. This principle is incorporated in the AFLP technique (EP-A 0 534 858). Using AFLP, the primer-binding sites result from a digestion of the target nucleic acid (i.e. total genomic DNA or cDNA) with one or more restriction enzymes, followed by adapter ligation. AFLP essentially targets a random selection of sequences contained in the target nucleic acid. It has been shown that, using AFLP, a practically unlimited number of target sequences can be amplified in a single reaction, depending on the number of target sequences that contain primer-binding region(s) that are perfectly complementary to the amplification primers. Exploiting the use of single primer-pair for amplification in combination with a non-random method for SNP target selection and efficient use of a high throughput detection platform may therefore substantially increase the efficiency of SNP genotyping, however such technology has not been provided in the art yet.

[0006] One of the principal methods used for the analysis of the nucleic acids of a known sequence is based on annealing two probes to a target sequence and, when the probes are hybridised adjacently to the target sequence, ligating the probes. The OLA-principle (Oligonucleotide Ligation Assay) has been described, amongst others, in US 4,988,617 (Landegren *et al.*). This publication discloses a method for determining the nucleic acid sequence in a region of a known nucleic acid sequence having a known possible mutation. To detect the mutation, oligonucleotides are selected to anneal to immediately adjacent segments of the sequence to be determined. One of the selected oligonucleotide probes has an end region wherein one of the end region nucleotides is complementary to either the normal or to the mutated nucleotide at the corresponding position in the known nucleic acid sequence. A ligase is provided which covalently connects the two probes when they are correctly base paired and are located immediately adjacent

to each other. The presence or absence of the linked probes is an indication of the presence of the known sequence and/or mutation.

[0007] US 5,876,924 by Zhang *et al.* also describes a ligation reaction using two adjacent probes wherein one of the probes is a capture probe with a binding element such as biotin. After ligation, the unligated probes are removed and the ligated captured probe is detected using paramagnetic beads with a ligand (biotin) binding moiety.

[0008] Abbot *et al.* in WO 96/15271 developed a method for a multiplex ligation amplification procedure comprising of the hybridisation and ligation of adjacent probes. These probes are provided with an additional length segment, the sequence of which, according to Abbot *et al.*, is unimportant. The deliberate introduction of length differences intends to facilitate the discrimination on the basis of fragment length in gel-based techniques.

[0009] WO 97/45559 (Barany *et al.)* describes a method for the detection of nucleic acid sequence differences by using combinations of ligase detection reactions (LDR) and polymerase chain reactions (PCR). Disclosed are methods comprising annealing allele-specific probe sets to a target sequence and subsequent ligation with a thermostable ligase, optionally followed by removal of the unligated primers with an exonuclease. Amplification of the ligated products with fluorescently labelled primers results in a fluorescently labelled amplified product. Detection of the products is based on separation by size or electrophoretic mobility or on an addressable array.

[0010] Detection of the amplified probes is performed on a universally addressable array containing capturing oligonucleotides. These capturing oligonucleotides contain a region that is capable of annealing to a pre-determined region in the amplified probe, a so-called zip-region or zip code. Each amplified probe contains a different zip code and each zip code will hybridise to its corresponding capturing oligonucleotide on the array. Detection of the label in combination with the position on the array provides information on the presence of the target sequence in the sample. This method allows for the detection of a number of nucleic acid sequences in a sample. However, the design, validation and routine use of arrays for the detection of amplified probes involves many steps (ligation, amplification, optionally purification of the amplified material, array production, hybridisation, washing, scanning and data quantification), of which some (particularly hybridisation and washing) are difficult to automate. Array-based detection is therefore laborious and costly to analyse a large number of samples for a large number of SNPs.

[0011] The LDR oligonucleotide probes in a given set may generate a unique length product and thus may be distinguished from other products based on size. For the amplification a primer set is provided wherein one of the primers contains a label. Different primers can be provided with different labels to allow for the distinction of products.

[0012] The method and the various embodiments described by Barany *et al.* are found to have certain disadvantages. One of the major disadvantages is that the method in principle does not provide for a true high throughput process for the determination of large numbers of target sequences in short periods of time using reliable and robust methods without compromising the quality of the data produced and the efficiency of the process.

[0013] More in particular, one of the disadvantages of the means and methods as disclosed by Barany *et al.* resides in the limited multiplex capacity when discrimination is based *inter alia*, on the length of the allele specific probe sets. Discrimination between sequences that are distinguishable by only a relatively small length difference is, in general, not straightforward and carefully optimised conditions may be required in order to come to the desired resolving power. Discrimination between sequences that have a larger length differentiation is in general easier to accomplish. This may provide for an increase in the number of sequences that can be analysed in the same sample. However, providing for the necessary longer nucleotide probes is a further hurdle to be taken. In the art, synthetic nucleotide sequences are produced by conventional chemical step-by-step oligonucleotide synthesis with a yield of about 98.5% per added nucleotide. When longer probes are synthesised (longer than ca. 60 nucleotides) the yield generally drops and the reliability and purity of the synthetically produced sequence can become a problem.

[0014] These and other disadvantages of the methods disclosed in WO 97/45559 lead the present inventors to the conclusion that the methods described therein are less preferable for adaptation in a high throughput protocol that is capable of handling a large number of samples each comprising large numbers of sequences.

[0015] The specific problem of providing for longer probes has been solved by Schouten *et al.* (WO 01/61033). WO 01/61033 discloses the preparation of longer probes for use in ligation-amplification assays. They provided probes that are considerably longer than those that can be obtained by conventional chemical synthesis methods to avoid the problem associated with the length-based discrimination of amplified products using slab-gels or capillary electrophoresis, namely that only a small part of the detection window / resolving capacity of up to 1 kilo base length is used when OLA probes are synthesised by chemical means. With an upper limit in practice of around 100-150 bases for chemically synthesised oligonucleotides according to the current state of technology, this results in amplification products that are less than 300 base pairs long at most, but often much less (see Barany *et al.)*. The difficulty of generating such long probes (more than about 150 nucleotides) with sufficient purity and yield by chemical means has been countered by Schouten *et al.*, using a method in which the probes have been obtained by an in vivo enzymatic template directed polymerisation, for instance by the action of a DNA polymerase in a suitable cell, such as an M13 phage.

[0016] However, the production and purification of such biological probes requires a collection of suitable host strains containing M13 phage conferring the desired length variations and the use of multiple short chemically synthesised

oligonucleotides in the process, thus their use is very laborious and time-consuming, hence costly and not suitable for high-throughput assay development. Furthermore, the use of relatively long probes and relatively large length differences between the amplifiable target sequences may result in differential amplification efficiencies in favour of the shorter target sequences. This adversely affects the overall data quality, hampering the development of a true high throughput method. Thus the need for a reliable and cost-efficient solution to multiplex amplification and subsequent length-based detection for high throughput application remains.

[0017] Other problems in the art and solutions provided thereto by the present invention will become clear throughout the description, the figures and the various embodiments and examples.

Description of the invention

[0018] The present invention relates to methods for high throughput separation and detection of multiple sequences. The present method resolves many of the problems previously encountered in the art. More in particular the present invention provides for a multiple ligation and amplification assay in combination with a multiple injection-multiple channel routine that allows for the rapid and high throughput analysis of a multiplicity of samples, preferably containing a multiplicity of sequences, in one channel and/or the simultaneous analysis of multiple samples over multiple channels and/or the simultaneous analysis of multiple samples over multiple channels carried out consecutively. The present invention also provides for a method for the high throughput discrimination and detection of a multitude of nucleotide sequences based on a combination of length differences and labels.

Detailed description of the invention

[0019] In a first aspect the invention relates to a method for high throughput separation and detection of a multiplicity of target sequence in a multiplicity of samples comprising subjecting each sample to a ligation-dependent amplification assay followed by a multiple injection step comprising consecutive and/or simultaneous injection of a multiplicity of samples.

[0020] The method preferably is a method for determining the presence or absence of at least two (different) target sequences (2) in at least two nucleic acid samples. The method preferably comprises the steps of:

(a) providing to a nucleic acid sample a pair of a first and a second oligonucleotide probe for each target sequence to be detected in the sample, whereby the first oligonucleotide probe has a section (4) at its 5'-end that is complementary to a first part (5) of a target sequence and the second oligonucleotide probe has a section (6) at its 3'-end that is complementary to a second part (7) of the target sequence, whereby the first (5) and second part (7) of the target sequence are located adjacent to each other, and whereby the first and second oligonucleotide probes (4, 6) each comprise a tag sequence (8, 9), whereby the tag sequences are essentially non-complementary to the target sequence, whereby the tag sequences may comprise a stuffer sequence (10, 11) and whereby the tag sequences comprise primer-binding sequences (12, 13);

(b) allowing the oligonucleotide probes to anneal to the adjacent parts of target sequences whereby the complementary sections (4,6) of the first and second oligonucleotide probes are adjacent;

(c) providing means (14) for connecting the first and the second oligonucleotide probes annealed adjacently to the target sequence and allowing the complementary sections (4,6) of the adjacently annealed first and second oligonucleotide probes to be connected, to produce a connected probe (15) corresponding to a target sequence in the sample;

(d) amplifying the connected probes from a primer pair comprising a first primer (16) that is complementary to the primer-binding sequence (12) of the first oligonucleotide probe (4) and a second primer (17) that is complementary to the primer-binding sequence (13) of the second oligonucleotide probe, to produce an amplified sample (19) comprising amplified connected probes (20);

(e) repeating steps (a) to (d) to generate at least two amplified samples (19);

(f) consecutively applying at least part the amplified samples (19) obtained in steps (d) and (e), to an application location of a channel (21) of an electrophoretic device (22), electrophoretically separating the amplified connected probes in the amplified samples (19) and detecting the separated amplified connected probes at a detection location (24) located distal from the application location of the channel; whereby the time period (23) between the consecutively applied amplified samples is such that the slowest migrating amplified connected probe (19) in an amplified sample is detected at the detection location(24), before the fastest migrating amplified connected probe of a subsequently applied amplified sample is detected at the detection location (24);

(g) determining the presence or absence of a target sequence in a sample by detecting the presence or absence of the corresponding connected probe.

[0021] In step (a) a multiplicity of different target sequences, i.e. at least two different target sequences, is brought into contact with a multiplicity of specific oligonucleotide probes under hybridising conditions. The pairs of oligonucleotide probes are subsequently allowed to anneal to the adjacent complementary parts of the multiple target sequences in the sample. Methods and conditions for specific annealing of oligonucleotide probes to complementary target sequences are well known in the art (see e.g. in Sambrook and Russel (2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press). Usually, after mixing of the oligonucleotide probes and target sequences the nucleic acids are denatured by incubation for a short period of time (e.g. 30 seconds to 5 minutes) in a low salt buffer (e.g. a buffer containing no salts or less salts than the ionic strength equivalent of 10 mM NaCl). The sample containing the denatured probes and target sequences is then allowed to cool to an optimal hybridisation temperature for specific annealing of the probes and target sequences, which usually is about 5°C below the melting temperature of the hybrid between the complementary section of the probe and its complementary sequence (in the target sequence). In order to prevent aspecific or inefficient hybridisation of one of the two probes in a primer pair, or in a sample with multiple target sequences, it is preferred that, within one sample, the sections of the probes that are complementary to the target sequences are of a similar, preferably identical melting temperatures between the different target sequences present in the sample. Thus, the complementary sections of the first and second probes preferably differ less than 20, 15, 10, 5, or 2 °C in melting temperature. This is facilitated by using complementary sections of the first and second probes with a similar length and similar G/C content. Thus, the complementary sections preferably differ less than 20, 15, 10, 5, or 2 nucleotides in length and their G/C contents differ by less than 30, 20, 15, 10, or 5 %. Complementary as used herein means that a first nucleotide sequence is capable of specifically hybridising to second nucleotide sequence under normal stringency conditions. A nucleotide sequence that is considered complementary to another nucleotide sequence may contain a minor amount, i.e. preferably less than 20, 15, 10, 5 or 2%, of mismatches. Alternatively, it may be necessary to compensate for mismatches e.g. by incorporation of so-called universal nucleotides, such as for instance described in EP-A 974 672, incorporated herein by reference. Since annealing of probes to target sequences is concentration dependent, annealing is preferably performed in a small volume, i.e. less than 10 µl. Under these hybridisation conditions, annealing of probes to target sequences usually is fast and does not to proceed for more than 5, 10 or 15 minutes, although longer annealing time may be used as long as the hybridisation temperature is maintained to avoid aspecific annealing. To avoid evaporation during denaturation and annealing, the walls and lids of the reaction chambers (i.e. tubes or microtitre wells) may also be heated to the same temperature as the reaction mixture. In preferred oligonucleotide probes the length of the complementary section is preferably at least 15, 18 or 20 nucleotides and preferably not more than 30, 40, or 50 nucleotides and the probes preferably have a melting temperature of at least 50°C, 55°C or 60°C.

[0022] In addition to the above hybridisation criteria, the complementary sections of the oligonucleotide probes are designed such that for each target sequence in a sample, a pair of a first and a second probe is provided, whereby the probes each contain a section at their extreme ends that is complementary to a part of the target sequence and the corresponding complementary parts of the target sequence are located essentially adjacent to each other. Within a pair of oligonucleotide probes, the first oligonucleotide probe has a section (4) at its 5'-end that is complementary to a first part (5) of a target sequence and the second oligonucleotide probe has a section (6) at its 3'-end that is complementary to a second part (7) of the target sequence. Thus, when the pair of probes is annealed to complementary parts of a target sequence the 5'-end of the first oligonucleotide probe is essentially adjacent to the 3'-end of the second oligonucleotide probe such that the respective ends of the two probes may be ligated to form a phosphodiester bond.

[0023] The respective 5'- and 3'-ends of a pair of first and second oligonucleotide probes that are annealed essentially adjacent to the complementary parts of a target sequence are connected in step (c) to form a covalent bond by any suitable means known in the art. The ends of the probes may enzymatically connected in a phosphodiester bond by a ligase, preferably a DNA ligase. DNA ligases are enzymes capable of catalysing the formation of a phosphodiester bond between (the ends of) two polynucleotide strands bound at adjacent sites on a complementary strand. DNA ligases usually require ATP (EC 6.5.1.1) or NAD (EC 6.5.1.2) as a cofactor to seal nicks in double stranded DNA. Suitable DNA ligase for use in the present invention are T4 DNA ligase, *E. coli* DNA ligase or preferably a thermostable ligase like e.g. *Thermus aquaticus* (Taq) ligase, *Thermus thermophilus* DNA ligase, or *Pyrococcus* DNA ligase. Alternatively, chemical autoligation of modified polynucleotide ends may be used to ligate two oligonucleotide probes annealed at adjacent sites on the complementary parts of a target sequence (Xu and Kool, 1999, Nucleic Acid Res. 27: 875-881).

[0024] Both chemical and enzymatic ligation occur much more efficient on perfectly matched probe-target sequence complexes compared to complexes in which one or both of the probes form a mismatch with the target sequence at, or close to the ligation site (Wu and Wallace, 1989, Gene 76: 245-254; Xu and Kool, *supra).* In order to increase the ligation specificity, i.e. the relative ligation efficiencies of perfectly matched oligonucleotides compared to mismatched oligonucleotides, the ligation is preferably performed at elevated temperatures. Thus, in a preferred embodiment of the invention, a DNA ligase is employed that remains active at 50 - 65°C for prolonged times, but which is easily inactivated at higher temperatures, e.g. used in the denaturation step during a PCR, usually 90 - 100°C. One such

DNA ligase is a NAD requiring DNA ligase from a Gram-positive bacterium (strain MRCH 065) as known from WO 01/61033. This ligase is referred to as "Ligase 65" and is commercially available from MRC Holland, Amsterdam.

[0025] A preferred method of the invention further comprises a step for the removal of oligonucleotide probes that are not annealed to target sequences and/or that are non-connected/ligated. Removal of such probes preferably is carried out prior to amplification, and preferably by digestion with exonucleases.

[0026] By removal/elimination of the oligonucleotide probes that are not connected/ligated a significant reduction of ligation independent (incorrect) target amplification can be achieved, resulting in an increased signal-to-noise ratio. One solution to eliminate one or more of the non-connected/ligated components without removing the information content of the connected probes is to use exonuclease to digest non-connected/ligated oligonucleotide probes. By blocking the end that is not ligated, for example the 3' end of the downstream oligonucleotide probe, one probe can be made substantially resistant to digestion, while the other is sensitive. Only the presence of a full-length ligation product sequence will then prevent digestion of the connected probe. Blocking groups include use of a thiophosphate group and/or use of 2-O-methyl ribose sugar groups in the backbone. Exonucleases include Exo I (3'-5'), Exo III (3'-5'), and Exo IV (both 5'-3' and 3'-5'), the later requiring blocking on both sides. One convenient way to block both probes is by using one long "padlock" probe (see M. Nilsson et. al., "Padlock Probes: Circularizing Oligonucleotides for Localized DNA Detection," Science 265: 2085-88 (1994), which is hereby incorporated by reference), although this is by no means required.

[0027] An advantage of using exonucleases, for example a combination of Exo I (single strand specific) and Exo III (double strand specific), is the ability to destroy both target and one of the oligonucleotide probes, while leaving the ligation product sequences substantially undigested. By using an exonuclease treatment prior to amplification, either one or both oligonucleotide probes in each set are substantially reduced, and thus hybridisation of the remaining oligonucleotide probes to the original target DNA (which is also substantially reduced by exonuclease treatment) and formation of a ligation product sequence which is a suitable substrate for PCR amplification by the oligonucleotide primer set is substantially reduced.

[0028] The oligonucleotide probes further contain a tag that is essentially non-complementary to the target sequence. The tag does not or not significantly hybridise, preferably at least not under the above annealing conditions, to any of the target sequences in a sample, preferably not to any of the sequences in a sample. The tag preferably comprises a primer-binding site and may optionally comprise a stuffer sequence of variable length (see below). The connected probes are amplified using a pair of primers corresponding to the primer-binding sites. In a preferred embodiment at least one of the primers or the same set of primers is used for the amplification of two or more different connected probes in a sample, preferably for the amplification of all connected probes in a sample. The different primers that are used in the amplification in step (d) are preferably essentially equal in annealing and priming efficiency. Thus, the primers in a sample preferably differ less than 20, 15, 10, 5, or 2 °C in melting temperature. This can be achieved as outlined above for the complementary section of the oligonucleotide probes. Unlike the sequence of the complementary sections, the sequence of the primers is not dictated by the target sequence. Primer sequences may therefore conveniently be designed by assembling the sequence from tetramers of nucleotides wherein each tetramer contains one A,T,C and G or by other ways that ensure that the G/C content and melting temperature of the primers are identical or very similar. The length of the primers (and corresponding primer-binding sites in the tags of the probes) is preferably at least 12, 15 or 17 nucleotides and preferably not more than 25, 30, 40 nucleotides.

[0029] In step (d) of the method of the invention, the connected probes are amplified to produce a (detectable) amplified connected probes by any suitable nucleic acid amplification method known in the art. Nucleic acid amplification methods usually employ two primers, dNTPs, and a (DNA) polymerase. A preferred method for amplification is PCR. "PCR" or "Polymerase Chain Reaction" is a rapid procedure for in vitro enzymatic amplification of a specific DNA segment. The DNA to be amplified is denatured by heating the sample. In the presence of DNA polymerase and excess deoxynucleotide triphosphates, oligonucleotides that hybridise specifically to the target sequence prime new DNA synthesis. One round of synthesis results in new strands of indeterminate length, which, like the parental strands, can hybridise to the primers upon denaturation and annealing. The second cycle of denaturation, annealing and synthesis produces two single-stranded products that together compose a discrete double-stranded product, exactly the length between the primer ends. This discrete product accumulates exponentially with each successive round of amplification. Over the course of about 20 to 30 cycles, many million-fold amplification of the discrete fragment can be achieved. PCR protocols are well known in the art, and are described in standard laboratory textbooks, e.g. Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley & Sons, Inc. (1995). Suitable conditions for the application of PCR in the method of the invention are described in EP-A 0 534 858 and Vos *et al.* (1995; Nucleic Acids Res. 23: 4407-4414), where multiple DNA fragments between 70 and 700 nucleotides and containing identical primer-binding sequences are amplified with near equal efficiency using one primer pair. Other multiplex and/or isothermal amplification methods that may be applied include e.g. LCR, self-sustained sequence replication (3SR), Q-β-replicase mediated RNA amplification, rolling circle amplification (RCA) or strand displacement amplification (SDA). In some instances this may require replacing the primer-binding sites in the tags of the probes by a suitable (RNA) polymerase-binding site.

[0030] Steps (a) to (d) may be performed on two or more nucleic acid samples, each containing two or more different target nucleic acids, to produce two or more amplified samples in which is presence or absence of connected and amplified probes is analysed. The amplified connected probes in a sample are preferably analysed on an electrophoretic device. The electrophoretic device preferably separates the different amplified connected probes in an amplified sample on the basis of length, after which the separated amplified connected probes may be detected as described below. A suitable electrophoretic device may be a gel-electrophoresis device, e.g. for conventional (polyacrylamide) slab gel-electrophoresis, or a capillary electrophoresis device such as exemplified by the MegaBACE equipment available from Molecular Dynamics Amersham-Pharmacia. The electrophoretic device preferably is a multichannel device in which multiple samples are electrophoresed in multiple channels in parallel. The electrophoretic device has an application location (per channel) for application (loading) of the amplified sample to be electrophoresed, a separation area over which the fragments in the sample migrate by electrophoresis, and preferably also a detection device located at a detection location distal from the application location. The detection device will usually comprises a photomultiplier for the detection of fluorescence, phosphorescence or chemiluminescence. Alternatively, in the case of gel-electrophoresis, the separated fragments may be detected in the gel e.g. by autoradiography or fluorography.

[0031] Preferably in the method of the invention, (parts of) two or more different amplified samples are applied consecutively to the same channel of the electrophoretic device. Depending on the electrophoresis conditions, the time period (23) between two (or more) consecutively applied amplified samples is such that the slowest migrating amplified connected probe (19) in an amplified sample is detected at the detection location(24), before the fastest migrating amplified connected probe of a subsequently applied amplified sample is detected at the detection location (24). Thus, the time intervals between subsequent multiple injections in one channel of the device are chosen such that consecutively applied samples after separation do not overlap at a point of detection.

[0032] The method according to the invention allows for the high throughput analysis of a multiplicity of samples each comprising a multiplicity of different target sequences by the consecutive injection of amplified samples, comprising amplified connected probes corresponding to the target sequences in the samples, in a channel of a multichannel electrophoretic device such as a capillary electrophoresis device. The method according to the invention allows for the analysis of a multiplicity of target sequences in a multiplicity of samples on a multiplicity of channels, thereby significantly increasing the throughput of the number of samples that can be analysed in a given time frame compared to conventional methods for the analysis of nucleotide sequences. "Throughput" as used herein, defines a relative parameter indicating the number of samples and target sequences that can be analysed per unit of time.

[0033] In the nucleic acid sample, the nucleic acids comprising the target may be any nucleic acid of interest. Even though the nucleic acids in the sample will usually be in the form of DNA, the nucleotide sequence information contained in the sample may be from any source of nucleic acids, including e.g. RNA, polyA$^+$ RNA, cDNA, genomic DNA, organellar DNA such as mitochondrial or chloroplast DNA, synthetic nucleic acids, DNA libraries, clone banks or any selection or combinations thereof. The DNA in the nucleic acid sample may be double stranded, single stranded, and double stranded DNA denatured into single stranded DNA. Denaturation of double stranded sequences yields two single stranded fragments one or both of which can be analysed by probes specific for the respective strands. Preferred nucleic acid samples comprise target sequences on cDNA, genomic DNA, restriction fragments, adapter-ligated restriction fragments, amplified adapter-ligated restriction fragments. AFLP fragments or fragments obtained in an AFLP-template preamplification.

[0034] In its widest definition, the target sequence may be any nucleotide sequence of interest. The target sequence preferably is a nucleotide sequence that contains, represents or is associated with a polymorphism. The term polymorphism herein refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wild type form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms. A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations). A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele. Other polymorphisms include small deletions or insertions of several nucleotides, referred to as indels. A preferred target sequence is a target sequence that is associated with an AFLP® marker, i.e. a polymorphism that is detectable with AFLP®.

[0035] It is preferred that a sample contains two or more different target sequences, i.e. two or more refers to the identity rather than the quantity of the target sequences in the sample. In particular, the sample comprises at least two different target sequence, in particular at least 10, preferably at least 25, more preferably at least 50, more in particular at least 100, preferably at least 250, more preferably at least 500 and most preferably at least 1000 additional target sequences. In practice, the number of target sequences is limited, among others, by the number of connected probes. E.g., too many different pairs of first and second oligonucleotide probes in a sample may corrupt the reliability of the multiplex amplification step.

[0036] A further limitation is formed e.g. by the number of fragments in a sample that can be resolved by the electrophoretic device in one injection. The number can also be limited by the genome size of the organism or the transcriptome complexity of a particular cell type from which the DNA or cDNA sample, respectively, is derived.

[0037] For each target sequence for which the presence or absence in a sample is to be determined, a specific pair of first and second oligonucleotide probes is designed with sections complementary to the adjacent complementary parts of each target sequence as described above. Thus, in the method of the invention, for each target sequence that is present in a sample, a corresponding (specific) amplified connected probe may be obtained in the amplified sample. Preferably, a multiplicity of first and second oligonucleotide probes complementary to a multiplicity of target sequences in a sample is provided. A pair of first and second oligonucleotide probes for a given target sequence in a sample will at least differ in nucleotide sequence from probe pairs for other target sequences, and will preferably also differ in length from probe pairs for other targets, more preferably a probe pair for a given target will produce a connected probe and/or amplified connected probe that differs in length from connected probes corresponding to other targets in the sample as described below. Alternatively, amplified connected probes corresponding to different targets may have an identical length if they can be otherwise distinguished e.g. by different labels as described below.

[0038] The oligonucleotide probes used in the present invention can also be circular probes. When using circular probes, a single linear oligonucleotide probe is provided in step (a) each target sequence in a sample. The single linear oligonucleotide probe combines the first and second oligonucleotide probes into a single molecule that is circularised in step (c) when the annealed complementarity sections are connected. Thus, in the single linear probe the sections of target complementarity as described for the first and second oligonucleotide probes are each present at the extreme ends of the single linear probe. The complementarity sections at the extreme ends are intervened by the sequences that may serve as primer-binding sequences as described above for the first and second oligonucleotide probes and may further be intervened by stuffer sequences of variable length.. An example of such an arrangement of functional groups in the circular probe is: (target-complementarity section 1 - stuffer sequence 1, primer-binding sequence 1 - primer-binding sequence 2 - stuffer sequence 2 - target-complementarity section 2). The skilled person will appreciate that the circular probes are synthesised and applied in a linear form and that they will only be circular when the two complementary sections at the extreme ends of the probe are connected (ligated) annealing to the appropriate target sequence. Circular probes can be advantageous in the ligation step (c) because both target-complementarity sections are contained in the same molecule, hence equimolar amounts of first and second probes are likely to hybridise to their respective target sequences because hybridisation of the first target-complementarity section to the target facilitates hybridisation of the second one and *vice versa*. In addition, the use of circular probes reduces the chances of the formation of incorrect ligation products that result from ligation between probes of different target sequences, due to the lower number of possible combinations of ligation products that can be formed when the first and second probes are part of the same circular molecule.

[0039] The probes that are not complementary to a part of the target sequence or that contain too many mismatches will not or only to a reduced extent hybridise to the target sequence when the sample is submitted to hybridisation conditions. Accordingly ligation is less likely to occur. The number of spurious ligation products from these probes in general will therefore not be sufficient and much smaller than the *bona fide* ligation products such that they are outcompeted during subsequent multiplex amplification. Consequently, they will not be detected or only to a minor extent.

[0040] The tag of the oligonucleotide probes may further comprise a stuffer sequence of a variable length. The length of the stuffer varies from 0 to 500, preferably from 0 to 100, more preferably from 1 to 50. The length of the tag varies from 15 to 540, preferably from 18 to 140, more preferably from 20 to 75.

[0041] To discriminate between different target sequences in the sample preferably a difference in length of the respective corresponding amplified connected probes is used. By separating the amplified connected probes based on length, the presence of the corresponding target sequences in the sample can be determined. Accordingly, in a preferred embodiment of the present invention, the discrimination between amplified connected probes derived from different target sequences in a sample is based on a length difference between the respective amplified connected probes corresponding to different target sequences in a sample or amplified sample.

[0042] Preferably, the length difference is provided by the length of the stuffer sequence(s) in the oligonucleotide probes. By including in each oligonucleotide probe a stuffer of a pre-determined length, the length of each amplified connected probe in an amplified sample can be controlled such that an adequate discrimination based on length differences of the amplified connected probes obtained in step (d) is enabled. In a preferred embodiment of a probe

according to the invention, the stuffer is located between the probe's section complementary to the target sequence and the primer-binding sequence. As such, the total length of the stuffer is provided by the combination of the length of the stuffer in the first probe and the length of the stuffer in the second probe. Accordingly, in a preferred embodiment, both the first oligonucleotide probes and the second oligonucleotide probes comprise a stuffer, preferably in their respective tags. A graphic representation thereof can be found in Figure 2. The length differentiation between amplified connected probes obtained from target sequences in the sample is preferably chosen such that the amplified connected probes can be distinguished based on their length. This is accomplished by using stuffer sequences or combinations of stuffer sequences in the first and second probes, which (together) result in length differences that may be distinguished on electrophoretic devices. Thus, from the perspective of resolving power, the length differences between the different amplified connected probes, as may be caused by their stuffers, are as large as possible. However, for several other important considerations, as noted before, the length differences between the different amplified connected probes is preferably as small as possible: (1) the upper limit that exists in practice with respect to the length of chemically synthesised probes of about 100-150 bases at most; (2) the less efficient amplification of larger fragments, (3) the increased chances for differential amplification efficiencies of fragments with a large length variation; and (4) the use of multiple injections of detection samples on the detection device which works best with fragments in a narrow length range. Preferably the length differences between the sequences to be determined and provided by the stuffers is at least sufficient to allow discrimination between essentially all amplified connected probes. By definition, based on chemical, enzymatic and biological nucleic acid synthesis procedures, the minimal useable size difference between different amplified connected probes in an amplified sample is one base, and this size difference fits within the resolving power of most electrophoresis devices, especially in the lower size ranges. Thus based on the above it is preferred to use multiplex assays with amplification products with differ in length by a single base(pair). In a preferred embodiment, the length difference between different amplified connected probes in an amplified sample is at least two nucleotides. In a particularly preferred embodiment of the invention the amplified connected probes corresponding to different target sequences in a sample have a length difference of two nucleotides.

[0043] The connected probes obtained from the ligation of the adjacent first and second probes are amplified in step (d), using a primer set, usually consisting of a pair of primers for each of the connected probes in the sample. The primer pair comprises primers that are complementary to primer-binding sequences that are present in the connected probes, preferably at the respective 3' and 5' ends of the connected probes. A primer pair usually comprises a first and at least a second primer, but may consist of only a single primer that primes in both directions.

[0044] In a preferred embodiment, at least one of the first and second oligonucleotide probes that are complementary to at least two different target sequences in a sample comprise a tag sequence that comprises a primer-binding site that is complementary to a single primer sequence. Thus, preferably at least one of the first and second primer in a primer pair is used for the amplification of connected probes corresponding to at least two different target sequences in a sample, more preferably for the amplification of connected probes corresponding to all target sequences in a sample. Preferably only a single first primer is used and in some embodiments only a single first and a single second primer is used for amplification of all connected probes. Using common primers for amplification of multiple different fragments usually is advantageous for the efficiency of the amplification step.

[0045] In a preferred embodiment, at least one of the primers complementary to the primer-binding sites of the first and second oligonucleotide probes in the sample comprises a label, preferably the second primer comprises a label. The label may be chosen from amongst the fluorescent or phosphorescent dyes in the group consisting of FAM, TET, JOE, NED, HEX, (ET-)ROX, FITC, Cy2, Texas Red, TAMRA, Alexa fluor 488™, Bodipy™ FL, Rhodamine 123, R6G, Bodipy 530, Alexafluor™532 and IRDyes ™.

[0046] By using a primer set comprising differently labelled primers, the number of connected probes that can be discriminated in a sample and hence the number of target sequences in a sample can be doubled for each additional label. Thus, for each additional label that is used in a sample, the number of target sequences that can be analysed in a sample is doubled. The maximum number of labels that can be used in one sample in a high throughput method is governed mostly by the limitations in the detection capabilities of the available detection platforms. At present, one of the most frequently used platforms (MegaBACE, by Molecular Dynamics -Amersham-Pharmacia Ltd. allows the simultaneous detection of up to four fluorescent dyes, being FAM, JOE or HEX, NED and (ET-)ROX. However, alternative capillary electrophoresis instruments are also suitable, which includes ABI310, ABI3100, ABI3700 (Perkin-Elmer Corp.), CEQ2000 XL (Beckman Coulter) and others. Nonlimiting examples of slab-gel based electrophoresis devices include ABI377 (Perkin Elmer Corp.) and the global IR^2 automated DNA sequencing system, available from LI-COR, Lincoln, Nebraska, USA.

[0047] In a preferred embodiment of the method of the invention, at least two groups of pairs of first and second oligonucleotide probes are provided to a sample, whereby each group of oligonucleotide probes has tag sequences with at least one group specific primer-binding site. The connected probes of each group are amplified from a primer pair wherein at least one of the first and second primers is complementary to the group specific primer-binding site, and whereby at least one of the first and second primers of a group comprises a group specific label. In each group,

an amplified connected probe corresponding to a target sequence in the sample, differs in length from an amplified connected probe corresponding to a different target sequence in the sample. The group specific labels are preferably such that the detection device can distinguish between the different group specific labels. The length difference is preferably provided by the length of the stuffer sequence. Preferably in this embodiment of the method of the invention, a first part of the groups has amplified connected probes having an even number of nucleotides and a second part of the groups has amplified connected probes having an odd number of nucleotides. Preferably, the groups of connected amplified probes having an even number of nucleotides and the groups of connected amplified probes having an odd number of nucleotides are labelled with (fluorescent) labels, which have the least overlap in their emission spectra. Thus, two groups of amplified connected probes, each group having an odd number of nucleotides are labelled with labels which have the least overlap in their emission spectra. The same holds for two groups of amplified connected probes, each group having an even number of nucleotides. Two groups of amplified connected probes, one group having an odd number of nucleotides and the other group having an even number of nucleotides are labelled with labels that have a larger overlap in their emission spectra. The relative notions as used herein of 'the least overlap in their emission spectra' and ' have a larger overlap in their emission spectra' refer to a group of labels from which a selection of the labels can be made for use in the present invention. This group of labels may depend on the detection platform used to other factors such as those disclosed herein before.. In a particularly preferred embodiment of this method, a first and second groups of connected amplified probes having an even number of nucleotides are produced and a third and fourth group of connected amplified probes having an odd number of nucleotides are produced and whereby the first and second group are labelled with FAM and NED, respectively, and the third and fourth group are labelled with (ET-)ROX and either JOE or HEX, respectively; or *vice versa*, whereby the first and second group are labelled with (ET-)ROX and either JOE or HEX, respectively, and the third and fourth group are labelled with FAM and NED, respectively. Thus, in these embodiments, the fluorescent labels are chosen such that the groups of amplified connected probes that co-migrate, because they both contain fragments with either even or odd numbers of nucleotides, have labels which have the least overlap in their emission spectra, thereby avoiding as much as possible cross-talk in the detection of amplified connected probe in different groups (see also below).

[0048] In order to come to a high throughput method of a multiplex of samples, a number of samples are treated similar to thereby generate a multiplicity of amplified detection samples which can then be analysed on a multichannel device which is at least capable of detecting the labels and/or length differences. Suitable devices are described herein above.

[0049] The multiplex analysis of the amplified samples following the method of the invention comprises applying at least part of an amplified sample to an electrophoretic device for subsequent separation and detection. Preferably such an amplified sample contains, or is at least suspected to contain, amplified connected probes, which is an indication that a target sequence has hybridised with the provided oligonucleotide probes and that those probes were annealed adjacently on the complementary target sequence so that they where connected, i.e. ligated. Subsequently, an amplified sample is subjected to a separating step for a selected time period before a next amplified sample is submitted. The period of time between two consecutively loaded amplified samples can be determined experimentally prior to executing the method. This period of time is selected such that, given the characteristics of an amplified sample, especially the difference in length between the shortest and the longest amplified connected probes in an amplified sample, as well as other experimental factors such as gel (matrix) and/or buffer concentrations, ionic strength etc., the fragments in an amplified sample are separated to such extent at the detection location which is located at the opposite end (distal) from the application location where the sample was applied, that the different amplified connected probes in a sample may be individually detected. After applying the last amplified sample, the separation can be continued for an additional period of time to allow the amplified connected probes of the last sample to be separated and detected. The combination of the selected period of time between applying two consecutive samples and the optional additional time period is chosen such that at the detection location the different amplified connected probes in consecutively applied samples are separated such that they may be individually detected, despite the limited length variation that exists between the different amplified connected probes within a single sample. Thus overlapping migration patterns are prevented when samples containing fragments of varying length are consecutively applied (injected) on the electrophoretic device..

[0050] The sample can be supplied with a nucleotide fragment size standard comprising one or more nucleotide fragments of known length. Methods of preparing and using nucleotide size standards are well known in the art (see e.g. Sambrook and Russel, 2001, *supra*). Such a size standard forms the basis for appropriate sizing of the amplified connected probes in the sample, and hence, for the proper identification of the detected fragment. The size standard is preferably supplied with every sample and/or with every injection. A size standard preferably contains a variety of lengths that preferably spans the entire region of lengths to be analysed. In a particular embodiment of the invention, it is considered advantageously to add flanking size standards from which the sizes of the amplified connected probes can be derived by interpolation. A flanking size standard is a size standard that comprises at least two labelled oligonucleotide sequences of which preferably one has a length that is at least one base shorter than the shortest amplified connected probe and preferably one that is a least one base longer than the longest amplified connected probe to

allow interpolation and minimise the introduction of further length variation in the sample. A preferred flanking size standard contains one nucleotide that is one nucleotide shorter the shortest amplified connected probe and one that is a least one base longer than the longest amplified connected probe and is labelled with at least one dye that is identical to the label used for labelling the amplified connected probes contained in the sample [A convenient way to assemble a suitable size standard is by (custom) chemical synthesis of oligonucleotides of the appropriate lengths, which are end-labelled with a suitable label. The size standard is applied with every consecutively applied sample to serve as local size references to size the loaded sample fragments. The size standard may be applied in the same channel or lane of the electrophoretic device as the sample to be analysed, i.e. together with the sample, or may be applied in a parallel channel or lane of a multichannel/lane device. The flanking size standard can be labelled with any of the labels used in the method. If the size standard is applied in the same channel of the device, the fragments of the standard are preferably labelled with a label that can be distinguished from the labels used for the detection of the amplified connected probes in a sample.

[0051] The selected time period prevents that consecutively applied samples after separation have an overlap of connected probes at the detection point. The selected time period is influenced by i). the length of the amplified connected probes; ii). the length variation in the amplified connected probes; and iii). the detection device and its operating conditions. Applying samples and separating consecutively applied samples in the same channel can be repeatedly performed in one or more channels, preferably simultaneously to allow for consecutive electrophoretic separation of multiple samples in one channel and/or simultaneous analysis of multiple samples over multiple channels and/or simultaneous analysis of multiple samples over multiple channels carried out consecutively. A graphic representation thereof is given in Figure 8.

[0052] In itself, multiple injection has been described in WO 01/04618. This publication discloses an apparatus and a method for the increased throughput analysis of small compounds using multiple temporally spaced injections. The publication discloses that samples comprising primers, extended by one nucleotide (single nucleotide primer extension or SnuPE, also known as minisequencing) could be detected using multiple temporally spaced injections on a capillary electrophoresis device. Minisequencing is based on annealing a complementary primer to a previously amplified target sequence. Subsequent extension of the primer with a separately provided labelled nucleotide provides for identification of the nucleotide adjacent to the primer. Principally, the primer extension product is of a constant length. To increase throughput the use of successive injections of extension products of the same length per run is suggested. To further increase the throughput, primers of a different length can be used, varying typically from 15 to 25 nucleotides. In contrast, the present invention contemplates analysing multiplex amplification products themselves directly with a length variation typically between 50 and 150 nucleotides. This is significantly more economical than minisequencing or SnuPE as outlined hereinbefore because multiple target sequences are amplified in a single reaction, whereas with minisequencing or SnuPE amplification is carried out individually for each target sequence. Furthermore, the use of primers of a different length and complementary to the target sequence compromises the efficiency of the subsequent amplification step needed in the method of the present invention.

[0053] In a variant of the technology, the starting (DNA) material of multiple individuals are pooled such that less detection samples containing this material are loaded on the detection device, This can be advantageous in the case of Linkage Disequilibrium (LD mapping) when the objective is to identify amplified connected probes (such as those representing SNP alleles) that are specific for a particular pool of starting samples, for example pools of starting material derived from individuals which have different phenotypes for a particular trait.

[0054] Using the method according to the invention, it is in principle possible and preferred to continuously apply, load or inject samples. Preferably the device is able to perform such operation automatically, e.g. controlled by a programmable computer. Preferably the multichannel device is suitable for such operation or is at least equipped for a prolonged operation without maintenance such as replacement of buffers, parts etcetera. However, in practice this will generally not be the case. When a final sample is submitted it is generally needed to continue the separation for an additional time period until the last fragment of the final sample has been detected.

[0055] Detection of the labelled separated samples is performed by a detector to result in detection data. The detector is of course dependent on the general system on which the separation is carried out (capillary electrophoresis, slab-gel electrophoresis, fixed detector-continuous gel-electrophoresis) but is also depending on the label that is present on the primer, such as a fluorescent or a radioactive label. The label can be selected from a large group, amongst others comprising fluorescent and/or phosphorescent moieties such as dyes, chromophores, or enzymes, antigens, heavy metals, magnetic probes, phosphorescent moieties, radioactive labels, chemiluminescent moieties or electrochemical detecting moieties. Preferably the label is a fluorescent or phosphorescent dye, more preferably selected from the group of FAM, HEX, TET, JOE, NED, and (ET-)ROX. Dyes such as FITC, Cy2, Texas Red, TAMRA, Alexa fluor 488™, Bodipy™ FL, Rhodamine 123, R6G, Bodipy 530, Alexafluor™532 and IRDyes ™ by Licor as used on the NEN Glober IR^2 platform are also suitable for use in the present invention.

[0056] As discussed hereinbefore, throughput can be increased by the use of multiple labelled primers. One of the problems associated with the use of different labels in one sample is cross talk or residual cross talk. Cross talk or

residual cross talk, as used herein, refers to the overlap between the emission spectra of different (fluorescent) labels. For instance when fluorescent dyes are used, each dye has a different emission (and absorption) spectrum. In case of two dyes in one sample, these spectra can overlap and may cause a disturbance of the signal, which contravenes the quality of the data obtained. Particularly when two nucleotide fragments to be detected in a sample are labelled with a different label and one of the fragments is present in an abundant amount whereas the other is present only in minute amounts, residual cross talk can cause that the measured signal of the fragment that is present in only minute amounts is mostly derived from the emission of another label with an overlapping emission spectrum that is abundantly contained in a fragment with identical size of another sample. The reciprocal effect of the other dye may also occur but in this example its effect is probably less because of the abundance differences between the amplified connected probes labelled with the respective dyes.

[0057] Chehab *et al.* (Proc. Natl. Acad. Sci. USA, 86:9178-9182 (1989) have attempted to discriminate between alleles by attaching different fluorescent dyes to competing alleles in a single reaction tube by selecting combinations of labels such that the emission maximum of one dye essentially coincides with the emission minimum of the other dye. However, at a certain wavelength at which one dye expresses an absorption maximum, there is always also some remaining absorption from another dye present in the sample, especially when the sample contains multiple dyes.

[0058] This route to multiplex analysis was found to be limited in scale by the relatively few dyes that can be spectrally resolved. One of the major problems with the use of multiple dyes is that the emission spectra of different fluorescent labels often overlap. The resulting raw data signals have to be corrected for the contribution of similar size fragments that are detected simultaneously and are labelled with another fluorescent dye by a process called cross-talk correction. Cross-talk correction is commonly carried out by mathematical means, based on the known theoretical absorption spectra for both dyes, after "raw" data collection from the detection device. Mathematical correction is based on theoretical spectra and ignores that emission spectra of labels are sensitive and often affected by the composition of the detection sample. These sensitivities can affect the brightness and/or the wavelength of the emission. This means that parameters such as pH, temperature, excitation light intensity, non-covalent interactions, salt concentration and ionic strength strongly influence the resulting emission spectrum. In particular, it is known that the presence of residual salts in a sample affects the fluorescence signal emitted by the dye and is a critical factor in case of detection by capillary electrophoresis using electrokinetic injection because it then also affects the injection efficiency. Thus, spectral overlap is a potential source of error that negatively impacts on data quality in case of multiplex detection using different fluorescent dyes.

[0059] The present invention provides for a solution to this problem such that two (or more) labels with overlapping spectra can be used in the same sample without significantly affecting data quality. By a predetermined combination of length differences and labels, an increase in the number of target nucleotide sequences that can be detected in sample is obtained while the quality of the data remains at least constant. In a preferred embodiment of the invention, spectral overlap between two differently labelled sequences is reduced by the introduction of a length difference between the two sequences. This label-related length difference can be provided for by the length of the stuffer sequence as described above. The number of different labels that can be used in the same sample in the present method is at least two, preferably at least three, more preferably at least four. The maximum number of labels is functionally limited by the minimum of spectral overlap that remains acceptable, which for most applications typically amounts to less than 15 percent of the true signal, preferably less than 10 percent, more preferably lees than 5 percent and most preferably less than 1 percent of the true signal.

[0060] In order to avoid the potential influence of residual cross-talk on the data quality in case different samples are labelled with multiple fluorescent dyes with overlapping emission spectra and fragments with identical length are detected simultaneously in the same run, in a particular preferred embodiment it is preferred to choose the stuffer sequences such that amplified connected probes differ by at least two base pairs within a multiplex set and differ by a single base pair between multiplex sets labelled with the different dyes that have overlapping spectra. By doing so, the length of the fragments labelled with the respective dyes can be chosen such that the potential influence of residual cross-talk on the quality of the data is circumvented because unique combinations of fragments size and labelling dye are defined (Figure 3).

[0061] A particular preferred embodiment of the invention is directed to a method in which a sample comprising amplified connected probes is derived from a multiplicity of target sequences. These amplified connected probes are differently labelled, thereby defining groups of amplified connected probes carrying the same label. Within each group, the stuffer provided for a length difference of at least two, preferably two nucleotides. Between two groups with labels having spectral overlap, the stuffer provides a length difference of one nucleotide, effectively resulting in one group having an even number of nucleotides and one group having an odd number of nucleotides as described above.

[0062] In one aspect the present invention pertains to a method for the improved discrimination and detection of target sequences in a sample, comprising providing at least a two or more groups of oligonucleotide probes, wherein the amplified connected probes obtained with different groups of oligonucleotide probes have different labels, wherein substantially each amplified connected probe target sequence within a group has the same label, wherein within a

group of identically labelled amplified connected probes a length difference is provided between each identically labelled probe within that group, wherein between the first and second group an additional length difference is provided such that each amplified connected probe in the amplified sample is characterised by a combination of length of the sequence and the label.

[0063] In a preferred embodiment to avoid cross-talk it is therefore desirable to combine a difference in length with a different label when analysing a set of amplified connected probes in such a way that the influence of spectral overlap on the data quality is avoided by length differences between the amplified connected probes labelled with the dyes that have overlapping emission spectra.

[0064] It is preferred that in each sample the connected probes derived from each target sequence differ from any other connected probe in the sample in length, and/or in the label or, preferably in the combination of the length and the label. To provide for an adequate separation of the amplified connected probes of different length it is preferred that the length difference between two different connected probes is at least two nucleotides, preferably two. When detecting polymorphisms it is preferred that the difference in length between two or more (SNP) alleles of the polymorphism is not more than two, thereby ensuring that the efficiency of the amplification is similar between different alleles or forms of the same polymorphism. This implies that preferably both alleles are amplified with the same pair of primers and hence will be labelled with the same dye. In a preferred embodiment of the invention, the stuffers present in the tags of both the first and second oligonucleotide probes are used to provide the length differences (i.e. 0 to 500 nucleotides, bases or base pairs) between the amplified connected probes. The total length of the amplified connected probes and the variation in the length is governed mostly by the techniques by which these fragments are analysed. In the high throughput multiple injection method of the present invention, it is preferred that the range of lengths of amplified connected probes in an amplified sample has a lower limit of 40, 60, 80, or 100 and an upper limit of 120, 140, 160, or 180 nucleotides, bases or base pairs, for conventional (capillary) electrophoresis platforms. It is particularly preferred that the range of lengths of the amplified connected probes varies from 100 to 140 nucleotides. However, these number are strongly related to the current limits of the presently known techniques. Based on the knowledge provided by this invention, the skilled artisan is capable of adapting these parameters when other circumstances apply.

[0065] The reliability of the multiplex amplification is further improved by limiting the variation in the length of the amplified connected probes. Limitations in the length variation of amplified connected probes is preferred to use multiple injection more efficiently and further results in reduction of the preferential amplification of smaller amplified connected probes in a competitive amplification reaction with larger connected probes.. This improves the reliability of the high throughput method of the present invention. Together with the multiple injection protocol as herein disclosed, these measures, alone or in combination provide for a significant increase in throughput in comparison with the art. A further improvement of the high throughput capacity is obtained by limiting the number of different amplified connected probes in a sample. It is regarded as more efficient and economical to limit the multiplex capacity of the ligation/amplification step in combination with the introduction of a multiple injection protocol. One of the most advantageous aspects of the present invention lies in the combination of multiplex ligation, multiplex amplification, preferably with a single primer pair or with multiple primer pairs which each amplify multiple connected probes, repeated injection and multiplex detection of different labels. One of the further advantageous aspects of the present invention resides in the combined application of length differences with different (overlapping) labels such that each connected probe and hence each target sequence within one sample can be characterised by a unique combination of length and label. This allows for a significant improvement of the efficiency of the analysis of target sequences as well as a significant reduction in the costs for each target analysed.

[0066] One aspect of the invention pertains to the use of the method in a variety of applications. Application of the method according to the invention is found in, but not limited to, techniques such as genotyping, transcript profiling, genetic mapping, gene discovery, marker assisted selection, seed quality control, hybrid selection, QTL mapping, bulked segregant analysis, DNA fingerprinting and microsatellite analysis. Another aspect pertains to the simultaneous high throughput detection of the quantitative abundance of target nucleic acids sequences.

Detection of single nucleotide polymorphisms

[0067] One particular preferred application of the high throughput method according to the invention is found in the detection of single nucleotide polymorphisms (SNPs). A first oligonucleotide probe comprises a part that is complementary to a part of the target sequence that is preferably located adjacent to the polymorphic site, i.e. the single polymorphic nucleotide. A second oligonucleotide probe is complementary to the part of the target sequence such that its terminal base is located at the polymorphic site, i.e. is complementary to the single polymorphic nucleotide. If the terminal base is complementary to the nucleotide present at the polymorphic site in a target sequence, it will anneal to the target sequence and will result in the ligation of the two probes. When the end-nucleotide, i.e. the allele-specific nucleotide does not match, no ligation or only a low level of ligation will occur and the polymorphism will remain undetected.

[0068] When one of the target sequences in a sample is derived from or contains a single nucleotide polymorphism (SNP), in addition to the probes specific for that allele, further probes can be provided that not only allow for the identification of that allele, but also for the identification of each of the possible alleles of the SNP (co-dominant scoring). To this end a combination of types of probes can be provided: one type probe that is the same for all alleles concerned and one or more of the other type of probe which is specific for each of the possible alleles. These one or more other type of probes contain the same complementary sequence but differ in that each contains a nucleotide, preferably at the end, that corresponds to the specific allele. The allele specific probe can be provided in a number corresponding to the number of different alleles expected. The result is that one SNP can be characterised by the combination of one type of probe with four other type (allele-specific) probes, identifying all four theoretically possible alleles (one for A, T, C, and G), by incorporating stuffer sequences of different lengths (preferred) or different labels into the allele specific probes.

[0069] In a particular embodiment, preferably directed to the identification of single nucleotide polymorphisms, the first oligonucleotide probe is directed to a part of the target sequence that does not contain the polymorphic site and the second oligonucleotide probe contains, preferably at the end distal from the primer-binding sequence, one or more nucleotide(s) complementary to the polymorphic site of interest. After ligation of the adjacent probes, the connected probe is specific for one of the alleles of a single nucleotide polymorphism. The stuffer sequence contained in the first oligonucleotide probe is preferably indicative of the locus that is to be analysed. The stuffer sequence contained in the second probe is preferably indicative of the nucleotide complementary to the polymorphic site.

[0070] To identify the allele of polymorphic site in the target sequence, a set of oligonucleotide probes can be provided wherein one first probe is provided and one or more second probes. Each second probe then contains a specific nucleotide at the end of the complementary sequence, preferably the 3'-end, in combination with a known length of the stuffer. For instance, in case of an A/C polymorphism, the second probe can contain a specific nucleotide T in combination with a stuffer length of 2 nucleotides and another second probe for this polymorphism combines a specific nucleotide G with a stuffer length of 0. As the primers and the complementary parts of the probes are preferably the same length, this creates a length difference of the resulting amplified connected probes of 2 nucleotides. In case the presence and/or the absence of all four theoretically possible nucleotides of the polymorphic site is desired, the stuffer-specific nucleotide combination can be adapted accordingly. In this embodiment, it can be considered that the locus-specific information is coupled to the length of the stuffer in the first probe and the allele-specific information of the polymorphic site is coupled to the length of the second stuffer. The combined length of the two stuffers can then be seen as indicative of the locus-allele combination. In a sample containing multiple targets sequences, amplified with the same pair of amplification-primers (and hence label) or with multiple pairs of amplifications primers with labels that have overlapping emission spectra, the combined stuffer lengths are chosen such that all connected probes are of a unique length. In Figure 4 an illustration of this principle is provided of two loci and for each locus two alleles. In a preferred embodiment this principle can be extended to at least ten loci with at least two alleles per locus. A further advantage of using two stuffers, one in each probe, is that by incorporating the majority of the length of the stuffer in the first probe (i.e. the locus-specific probe) the allele-specific probes can remain shorter i.e. the minimum number of bases sufficient for discrimination between the allele specific probes, which saves costs. The incorporation of the complete stuffer sequence in the allele specific probe would require the synthesis of the majority of the stuffer sequence twice.

Detection of specific target sequence

[0071] The target sequence contains a known nucleotide sequence derived from a genome. Such a sequence does not necessarily contain a polymorphism, but is for instance specific for a gene, a promoter, an introgression segment or a transgene or contains information regarding a production trait, disease resistance, yield, hybrid vigour, is indicative of tumours or other diseases and/or gene function in humans, animals and plants. To this end, the complementary parts of the first probe and the second probe are designed to correspond to a, preferably unique, target sequence in genome, associated with the desired information. The complementary parts in the target sequence are located adjacent to each other. In case the desired target sequence is present in the sample, the two probes will anneal adjacently and after ligation and amplification can be detected.

Detection of AFLP markers

[0072] AFLP, its application and technology is described in Vos *et al.*, Nucleic Acids Research, vol. 23, (1995), 4407-4414 as well as in EP-A 0 534 858 and US 6045994, all incorporated herein by reference. For a further description of AFLP, its advantages, its embodiments, its techniques, enzymes, adapters, primers and further compounds, tools and definitions used, explicit reference is made to the relevant passages of the publications mentioned hereinbefore relating to AFLP. AFLP and its related technology is a powerful DNA fingerprinting technique for the identification of

for instance specific genetic markers (so-called AFLP-markers), which can be indicative of the presence of certain genes or genetic traits or can in general be used for comparing DNA, cDNA or RNA samples of known origin or restriction pattern. AFLP-markers are in general associated with the presence of polymorphic sites in a nucleotide sequence to be analysed. Such a polymorphism can be present in the restriction site, in the selective nucleotides, for instance in the form of indels or substitutions or in the rest of the restriction fragment, for instance in the form of indels or substitutions. Once an AFLP marker is identified as such, the polymorphism associated with the AFLP-marker can be identified and probes can be developed for use in the ligation assay of the present invention.

[0073] In another aspect the present invention pertains to a nucleic acid probe comprising a part that is capable of hybridising to part of a target sequence and further comprising a primer-binding sequence and a stuffer. The invention also pertains to a set of probes comprising of two or more probes wherein each probe comprises a part that is complementary to part of a target sequence and wherein the complementary parts of the probes are located essentially adjacent on the target sequence and wherein each probe further comprises a stuffer, which stuffer is located essentially next to the complementary part and a primer-binding sequence located essentially adjacent to the stuffer.

[0074] The invention in a further aspect, pertains to the use of a set of probes in the analysis of at least one nucleotide sequence and preferably in the detection of a single nucleotide polymorphism, wherein the set further comprises at least one additional probe that contains a nucleotide that is complementary to the known SNP allele. Preferably the set comprises a probe for each allele of a specific single nucleotide polymorphism. The use of a set of probes is further preferred in a method for the high throughput detection of single nucleotide polymorphisms wherein the length of the first stuffer in the first probe is specific for a locus of a single nucleotide polymorphism and the length or the presence of the second stuffer in the second probe is specific for an allele of the single nucleotide polymorphism.

[0075] Another aspect of the invention relates to the primers and more in particular to the set of primers comprising a first primer and one or more second primers, wherein each second primer contains a label and which second primer comprises a nucleotide sequence that is specific for said label.

[0076] The present invention also finds embodiments in the form of kits. Kits according to the invention are for instance kits comprising probes suitable for use in the method as well as a kit comprising primers, further a combination kit, comprising primers and probes, preferably all suitably equipped with enzymes buffers etcetera, is provided by the present invention.

[0077] The efficiency of the present invention can be illustrated as follows. When a capillary electrophoretic device with 96 channels and capable of detecting four labels simultaneously is used, allowing for 12 subsequent injections per run per channel with a empirically optimised minimum selected time period between the injections, a sample containing 20 target sequences of interest allows for the high throughput detection of 96 (channels) * 12 (injections) * 20 (targets) * 4 (labels) = 92160 target sequences, using the method of the present invention. In the case of co-dominant SNP-detection, data regarding 46080 SNPs can be detected in a single run.

Description of the Figures:

[0078]

Figure 1: Schematic representation of the oligonucleotide ligation-amplification assay, resulting in amplified connected probes.

Figure 2: Schematic representation of two connected probes, wherein

(a) only one probe contains a stuffer (10) and primer-binding sequences (12,13); and
(b) both probes contain a stuffer (10, 11) and primer-binding sequences (12,13).

Figure 3: Schematic representation of the unique combination of different lengths and labels with a schematic elution profile in one channel of a multichannel device.

Figure 4: Schematic representation of the oligonucleotide ligation-ligation assay of the present invention. The principle is represented for two loci 1 and 2 and for each locus two alleles for reasons of simplicity only, but can easily be extended to at least 10 loci with 2 alleles each. The primer set consists of one first primer (solid bold line) and one second primer (dashed bold line). The theoretically possible connected probes are schematically outlined, together with the primers. The connected probes differ in length.

Figure 5: Schematic representation of the oligonucleotide ligation-ligation assay of the present invention. The principle is represented for two loci 3 and 4 and for each locus two alleles. The primer set consists of one first primer and two second primers. The theoretically possible connected probes are schematically outlined, together with the primers. The connected probes differ in length and in label.

Figure 6: Schematic representation of the results of a sample containing 80 amplified connected probes with:

- a length difference between 135 base pairs (bp) to 97 bp for the amplified connected probes with an odd length and labelled with Label 1 and Label 3; and

- a length difference between 134 bp to 96 bp for the amplified connected probes with an even length and labelled with Label 2 and Label 4; and

- a flanking size ladder with oligonucleotides of 94/95 and 136/137 (bp) carrying label 1, 2, 3 or 4

Figure 7: Schematic representation of the separation profile in one channel, submitting one sample comprising multiple amplified connected probes labelled with Label 1, 2, 3, and 4. The multiple labelled amplified connected probes are detectably separated at the point of detection.

Figure 8: Schematic representation of the multiple injection of samples in one channel, with a graphic illustration of the selected time period (23) between the injection of subsequent samples and the additional time period (25) after submitting the last sample.

Figure 9: Schematic representation of the ligation of up to 40 loci, and the subsequent amplification and detection phase of the method. Depending on the complexity and the number of loci to be analysed, the points in the procedure at which pooling can be contemplated is indicated as an optional (dotted) feature). Amplification is here carried out by using one forward primer (Forward) and for each label one (differently labelled) reverse primer (Reverse 1, 2, 3, 4). When the ligation (sub)samples are pooled, there are in principle two options for amplification. For instance if (sub)samples derived from Loci 1-10 are pooled with (sub)samples derived from Loci 11-20 prior or subsequent to ligation, the pooled (sub)sample can be amplified with the Forward primer and the Reverse primers 1 and 2 in one step or in two steps, first with Forward and Reverse 1, followed by Forward and Reverse 2 or *vice versa*. Detection can also be performed in a similar way, detecting both labels simultaneously or first label 1, followed by label 2, optionally by double injection.

Figure 10: A gel of a multiplex oligonucleotide ligation assay of 12 SNPs from the Colombia ecotype, the Landsberg erecta ecotype and a 50/50 mixture of the Colombia and the Landsberg erecta ecotypes.

Figure 11: A. Partial electropherogram of FAM labelled detection of the Colombia sample on a capillary electrophoretic device (MEGABace). The same multiplex mixture was injected. Amplified connected probes in a size range 97-134 bp and flanking sizer fragments (designated S) are 94, 95 and 137 bp. Probes and sizers are all labelled with FAM.

B. Partial electropherogram of FAM labelled detection of the Landsberg erecta sample on a capillary electrophoretic device (MEGABace). The same multiplex mixture was injected. Amplified connected probes in a size range 97-134 bp and flanking sizer fragments (designated S) are 94, 95 and 137 bp. Probes and sizers are all labelled with FAM.

Figure 12: A: Raw trace file of a sample containing a 120 bp ET-ROX labelled fragment and a 124 bp NED -labelled fragments. Note the FAM and JOE labels from other labelled fragments in the sample with the same length. FAM and JOE have overlapping fluorescence spectra (ET-ROX and FAM, JOE and NED), resulting in overlapping signals (cross-talk) with sequences of equal length.

B: Mathematical cross-talk correction resulting in a processed, cross-talk corrected trace file. Cross talk is reduced, but remains of the overlapping spectra (FAM, JOE) are present, resulting in false positive (or negative) signals.

C, D, E, F: single label plots illustrate the presence of remnants (D, E) of the mathematical correction, compared to the positive signals (C, F)

Figure 13

A: Representation of the effect of incomplete removal of cross-talk of a 120 bp ET-ROC fragment and a 124 bp NED fragment, resulting in incorrect scored data, compared to theoretically expected data.

B: Representation of the effect of the use of cross-talk correction by length-label combinations. Scored data and expected data are correctly interpreted and false-positive or negative data are eliminated.

Examples

I. Design of the stuffer sequences

[0079] In order to prevent cross-hybridisation between the amplification products, it is preferred that the sequences of the stuffer sequences are different and do not form hairpins. In the tables 1-5, stuffer sequences are presented which can be used for the development of probes for each fluorescent dye, and have been verified for the absence of hairpins using Primer Designer version 2.0 (copyright 1990,1991, Scientific and Educational software) The stuffer sequences are assembled from randomly chosen tetramer blocks containing one G, C, T and A, and have therefore by definition

a 50% GC content. The stuffer sequence in the forward OLA probe for the two SNP alleles are kept identical to avoid preferential SNP allele amplification.

Table 1.

Lengths of stuffer sequences					
ET-ROX and JOE probes.			FAM and NED probes.		
Total stutter length	Stutter length 1st type probe	Stutter length 2nd type probe	Total stuffer length	Stuffer length 1st type probe	Stutter length 2nd type probe
0	0	0	1	1	0
2	0	2	3	1	2
4	4	0	5	5	0
6	4	2	7	5	2
8	8	0	9	9	0
10	8	2	11	9	2
12	12	0	13	13	0
14	12	2	15	13	2
16	16	0	17	17	0
18	16	2	19	17	2
20	20	0	21	21	0
22	20	2	23	21	2
24	24	0	25	25	0
26	24	2	27	25	2
28	28	0	29	29	0
30	28	2	31	29	2
32	32	0	33	33	0
34	32	2	35	33	2
36	36	0	37	37	0
38	36	2	39	37	2

Table 2. Stuffer sequences for ET-ROX probes (5'-3').

Stuffer length	
1st type probe	2nd type probe
0	0
0	2 CA
4 TGCA	0
4 TGCA	2 CA
8 ACGT TACG	0
8 ACGT TACG	2 CA
12 TAGC GTCA GCAT	0
12 TAGC GTCA GCAT	2 CA
16 CATG GCAT ACGT TACG	0
16 CATG GCAT ACGT TACG	2 CA
20 GATC GCTA ACGT TACG GCAT	0
20 GATC GCTA ACGT TACG GCAT	2 CA
24 TCGA GATC ACGT CATG CTGA GCAT	0
24 TCGA GATC ACGT CATG CTGA GCAT	2 CA
28 CAGT TCAG GCAT TCGA CTAG CGTA TACG	0
28 CAGT TCAG GCAT TCGA CTAG CGTA TACG	2 CA
32 GTCA ATCG GACT CTGA GACT CATG CGAT GACT	0
32 GTCA ATCG GACT CTGA GACT CATG CGAT GACT	2 CA
36 GATC CGAT CGAT ATCG ACGT AGCT GCAT CGTA ATCG	0
36 GATC CGAT CGAT ATCG ACGT AGCT GCAT CGTA ATCG	2 CA

Table 3. Stuffer sequences for JOE probes (5'-3').

Stuffer length	
First type probe	2nd type probe
0	0
0	2 TG
4 ACTG	0
4 ACTG	2 TG
8 GCAT CAGT	0
8 GCAT CAGT	2 TG
12 ATCG GCAT TACG	0
12 ATCG GCAT TACG	2 TG
16 TACG GCAT AGTC ACGT	0
16 TACG GCAT AGTC ACGT	2 TG
20 GATC GCTA ACGT TACG GCAT	0
20 GATC GCTA ACGT TACG GCAT	2 TG
24 CTAG ATGC TCAG GCTA TCGA CATG	0
24 CTAG ATGC TCAG GCTA TCGA CATG	2 TG
28 GTAC CGAT ACGT TAGC GACT TAGC CGTA	0
28 GTAC CGAT ACGT TAGC GACT TAGC CGTA	2 TG
32 CGTA ATCG GATC CGTA ACGT GCAT ATGC CAGT	0
32 CGTA ATCG GATC CGTA ACGT GCAT ATGC CAGT	2 TG
36 GACT TCGA GATC TGCA ACGT ACGT CGTA AGCT GCTA	0
36 GACT TCGA GATC TGCA ACGT ACGT CGTA AGCT GCTA	2 TG

Table 4. Stuffer sequences for FAM probes (5'-3').

Stuffer length	
First type probe	2nd type probe
1 C	0
1 C	2 GA
5 C GACT	0
5 C GACT	2 GA
9 C CGAT TAGC	0
9 C CGAT TAGC	2 GA
13 C ATCG GATC AGCT	0
13 C ATCG GATC AGCT	2 GA
17 C ATGC TAGC ACGT ACTG	0
17 C ATGC TAGC ACGT ACTG	2 GA
21 C GTAC CAGT CATG GATC CGAT	0
21 C GTAC CAGT CATG GATC CGAT	2 GA
25 C GATC ATCG ACTG GTAC TACG GACT	0
25 C GATC ATCG ACTG GTAC TACG GACT	2 GA
29 C GTAC GCAT GCTA ACGT TACG GACT ATCG	0
29 C GTAC GCAT GCTA ACGT TACG GACT ATCG	2 GA
33 C CGTA GCAT CGAT ATCG GTCA ACTG GATC AGCT	0
33 C CGTA GCAT CGAT ATCG GTCA ACTG GATC AGCT	2 GA
37 C GTAC CATG TCGA CGTA GATC CGTA TAGC ACTG AGTC	0
37 C GTAC CATG TCGA CGTA GATC CGTA TAGC ACTG AGTC	2 GA

Table 5. Stuffer sequences for NED probes (5'-3').

Stuffer length	
First type probe	2nd type probe
1 C	0
1 C	2 TC
5 C GTAC	0
5 C GTAC	2 TC
9 C GCAT TCGA	0
9 C GCAT TCGA	2 TC
13 C ATCG GCAT GACT	0
13 C ATCG GCAT GACT	2 TC
17 C GTCA ATGC ACGT TACG	0
17 C GTCA ATGC ACGT TACG	2 TC
21 C GCAT CGAT AGCT CTGA ACGT	0
21 C GCAT CGAT AGCT CTGA ACGT	2 TC
25 C GCAT ATCG GATC GATC GCAT ACGT	0
25 C GCAT ATCG GATC GATC GCTA ACGT	2 TC
29 C ATCG GATC CATG CGTA GCAT ATCG ACGT	0
29 C ATCG GATC CATG CGTA GCAT ATCG ACGT	2 TC
33 C TGCA AGTC CGAT TACG ATCG ACGT GCTA TGCA	0
33 C TGCA AGTC CGAT TACG ATCG ACGT GCTA TGCA	2 TC
37 C AGCT CAGT ATCG AGTC GACT ACGT TGCA TACG GATC	0
37 C AGCT CAGT ATCG AGTC GACT ACGT TGCA TACG GATC	2 TC

II. EXAMPLES MULTIPLEX LIGATION ASSAY AND DETECTION

Example 1. Description of biological materials and DNA isolation.

[0080] Recombinant Inbred (RI) lines generated from a cross between the *Arabidopsis* ecotypes Colombia and Landsberg *erecta* (Lister and Dean, Plant Journal, 4, pp 745-750, (1993) were used. Seeds from the parental and RI lines were obtained from the Nottingham Arabidopsis Stock Centre.

[0081] DNA was isolated from leaf material of individual seedlings using methods known *per se*, for instance essentially as described in EP-0534858, and stored in 1X TE (10 mM Tris-HCl pH 8.0 containing 1 mM EDTA) solution. Concentrations were determined by UV measurements in a spectrophotometer (MERK) using standard procedures, and adjusted to 100 ng / µl using 1X TE.

Example 2. Selection of Arabidopsis SNP's.

[0082] The Arabidopsis SNP's that were selected from *The Arabidopsis Information Resource* (TAIR) website: http://www.arabidopsis.org/SNPs.html:, are summarised in Table 6 in

Table 6.

Selected SNPs from *Arabidopsis thaliana.*			
	SNP	SNP alleles*	RI Map position
1	SGCSNP1	G/A	chr. 2; 72,81
2	SGCSNP20	A/C	chr. 4; 15,69
3	SGCSNP27	T/G	chr. 3; 74,81
4	SGCSNP37	C/G	chr 2; 72,45
5	SGCSNP39	T/C	chr. 5; 39,64
6	SGCSNP44	A/T	not mapped
7	SGCSNP55	C/A	chr. 5; 27,68
8	SGCSNP69	G/A	chr. 1; 81,84
9	SGCSNP119	A/T	chr. 4; 62,06
10	SGCSNP164	T/C	chr. 5; 83,73
11	SGCSNP209	C/G	chr. 1; 70,31
12	SGCSNP312	G/T	chr. 4; 55,95

* For all SNP's the allele preceding the backslash is the Colombia allele.

Example 3. Oligonucleotide probe design for oligonucleotide ligation reaction

[0083] The oligonucleotide probes (5'-3' orientation) were selected to discriminate the SNP alleles for each of the twelve SNP loci described in Example 2. PCR binding regions are underlined, stuffer sequences are double underlined. Reverse primers are phosphorylated at the 5' end:. p indicates phosphorylated. The sequences are summarised in Table 7.

TABLE 7 Oligonucleotide probes for detection of Colombia and Landsberg SNPs

SEQ.ID	Code	Nucleotide sequence
	SCGSNP1	
1	SNPfwd001 (G allele)	CGCCAGGGTTTTCCCAGTCACGACTTCAGGACTAGTCTATACCTTGAG
2	SNPfwd002 (A allele)	CGCCAGGGTTTTCCCAGTCACGACGACTTCAGGACTAGTCTATACCTTGAA
3	SNPrev001 (Common reverse SNP001)	pCTATGTGAACCAAATTAAAGTTTACTCCTGTGTGAAATTGTTATCCGCT
	SGCSNP20	
4	SNPfwd003 (A-allele)	CGCCAGGGTTTTCCCAGTCACGACCTGCTCTTTCCTCGCTAGCTTCAGA
5	SNPfwd004 (C-allele)	CGCCAGGGTTTTCCCAGTCACGACGACTGCTCTTTCCTCGCTAGCTTCAGC
6	SNPrev002 (common reverse SNP20):	pAGATTCGGACCTTCTCTCATAATCCGACTTCCTGTGTGAAATTGTTATCCGCT
	SGCSNP27	
7	SNPfwd005 (T-allele)	CGCCAGGGTTTTCCCAGTCACGACGAAGAGGAGAGTGGCTACGAACTCT
8	SNPfwd006 (G-allele)	CGCCAGGGTTTTCCCAGTCACGACGAGAAGAGGAGAGTGGCTACGAACTCG
9	SNPrev003 (common reverse SNP27)	pGCGATAACTGCTCTGTAGAAAGACCCGATTAGCTCCTGTGTGAAATTGTTATCCGCT
	SGCSNP37	
10	SNPfwd007 (C-allele)	CGCCAGGGTTTTCCCAGTCACGACAATCGGCCTAAGCAAGCTTGTTTTC
11	SNPfwd008 (G-allele)	CGCCAGGGTTTTCCCAGTCACGACGAAATCGGCCTAAGCAAGCTTGTTTTG
12	SNPrev004 (common reverse SNP37)	PTGCTATTGATATCTCTGTGCAACTCATCGGATCAGCTTCCTGTGTGAAATTGTTATCCGCT
	SGCSNP39	
13	SNPfwd009 (T-allele	CGCCAGGGTTTTCCCAGTCACGACGATCGGAAAGATATCGGAGCTCCTT

EP 1 319 718 A1

14	SNPfwd010 (C-allele)	CGCCAGGGTTTTCCCAGTCACGACGAGATCGGAAAGATATCGGAGCTCCTC
15	SNPrev005 (common reverse SNP39)	pGTCGGTGTCAACCGATCCACGGCGCATGCTAGCACGTACTGTCCTGTGTGAAATTG TTATCCGCT
	SGCSNP44	
16	SNPfwd011 (A-allele)	CGCCAGGGTTTTCCCAGTCACGACGAACTGGCATCAATCAGGCCTCCAA
17	SNPfwd012 (T-allele)	CGCCAGGGTTTTCCCAGTCACGACGAGAACTGGCATCAATCAGGCCTCCAT
18	SNPrev006 (common reverse SNP44)	pCCTTAATGCAAGGGCTTATTACGTCGTACCAGTCATGGATCCGATTCCTGTGTGAA ATTGTTATCCGCT
	SGCSNP55:	
19	SNPfwd013 (C-allele)	CGCCAGGGTTTTCCCAGTCACGACGGACTCCAAGGTATTGTTAGGCGCC
20	SNPfwd014 (A-allele)	CGCCAGGGTTTTCCCAGTCACGACGAGGACTCCAAGGTATTGTTAGGCGCA
21	SNPrev007 (common reverse SNP55)	pAACCACCAAGATCAGTCTCATCTTCGATCATCGACTGGTACTACGGACTTCCTGTG TGAAATTGTTATCCGCT
	SGCSNP69	
22	SNPfwd015 (G-allele)	CGCCAGGGTTTTCCCAGTCACGACCATCTCTTGCGCCTTCTCAGTGTTG
23	SNPfwd016 (A-allele)	CGCCAGGGTTTTCCCAGTCACGACGACATCTCTTGCGCCTTCTCAGTGTTA
24	SNPrev008 (common reverse SNP69)	pTGACGTCCGTCGAAGAATAGGTAACCGTACGCATGCTAACGTTACGGACTATCGTC CTGTGTGAAATTGTTATCCGCT
	SGCSNP119	
25	SNPfwd017 (A-allele)	CGCCAGGGTTTTCCCAGTCACGACAGTTTCAAAACCCATGACGCTTCTA
26	SNPfwd018 (T-allele)	CGCCAGGGTTTTCCCAGTCACGACGAAGTTTCAAAACCCATGACGCTTCTT
27	SNPrev009 (common reverse SNP119)	pGTGATAGCTGAAAAGACCCATTCTCCGTAGCATCGATATCGGTCAACTGGATCAG

	SGCSNP164	CTTCCTGTGTGAAATTGTTATCCGCT
28	SNPfwd019 (T-allele)	CGCCAGGGTTTTCCCAGTCACGACATACTCCAATTGCTCAGGCACAGTT
29	SNPfwd020 (C-allele)	CGCCAGGGTTTTCCCAGTCACGACGAATACTCCAATTGCTCAGGCACAGTC
30	SNPrev010 (common reverse SNP164)	pCTCCTTGTCCCACGAAGATAGTTCCGTACCATGTCGACGTAGATCCGTATAGCACT GAGTCTCCTGTGTGAAATTGTTATCCGCT
	SGCSNP209	
31	SNPfwd021 (C-allele)	CGCCAGGGTTTTCCCAGTCACGACGTAGAGGCTCTAAACAGCTGCTTCC
32	SNPfwd022 (G-allele).	CGCCAGGGTTTTCCCAGTCACGACGAGTAGAGGCTCTAAACAGCTGCTTCG
33	SNPrev011 (common reverse SNP209)	pCTTGTTTATGCTAAGGGCCGGCTCCTCCTGTGTGAAATTGTTATCCGCT
	SGCSNP312	
34	SNPfwd023 (G-allele)	CGCCAGGGTTTTCCCAGTCACGACTAAGTCAGCTCCTAAGCTTCCATCG
35	SNPfwd024 (T-allele)	CGCCAGGGTTTTCCCAGTCACGACGATAAGTCAGCTCCTAAGCTTCCATCT
36	SNPrev012 (common reverse SNP312)	pAAGCCACTTCCTCCTGCTCAAGCGCGACTTCCTGTGTGAAATTGTTATCCGCT

All oligonucleotides were purchased from MWG, Ebersberg, Germany. The concentration of the oligonucleotides was adjusted to 1 µM

Example 4. Design of the PCR amplification primers

[0084] The sequences of the primer used for PCR amplification were complementary to the PCR primer binding regions incorporated in the ligation probes described in Example 3. The sequences represent the so called M13 forward and M13 reverse primers. Usually the forward primer is labelled with FAM or α^{33}P-dATP depending on the detection platform. The sequence of the primers in 5'-3' orientation are:

M13 forward: CGCCAGGGTTTTCCCAGTCACGAC [SEQ ID No. 37]

M13 reverse: AGCGGATAACAATTTCACACAGGA [SEQ ID No.38]

[0085] The concentration of these oligo's was adjusted to 50 ng / µl.

Example 5. Buffers and Reagents

[0086] The composition of the buffers was: Hybridisation buffer (1X), 20 mM Tris-HCl pH 8.5, 5 mM $MgCl_2$, 100 mM KCl, 10 mM DTT, 1 mM NAD^+ Ligation buffer (1X) 20 mM Tris-HCl pH 7.6, 25 mM Kac, 10 mM $MgAc_2$, 10 mM DTT, 1 mM NAD^+ 0.1% Triton-X100.PCR buffer (10X):10x PCR buffer (contains 15 mM $MgCl_2$). (Qiagen, Valencia, United States of America)No additions were used in the PCR

Example 6. Ligation and Amplification

Ligation reactions:

[0087] Ligation reactions were carried out as follows: 100 ng genomic DNA (1 µl of 100 ng / µl) in 5 µl total volume was heat denatured by incubation for 5 minutes at 94 °C and cooled on ice. Next 4 fmol of each OLA forward and reverse probes described in Example 3 (36 oligonucleotides in total) were added, and the mixture was incubated for 16 hours at 60 °C. Next, 1 unit of Taq Ligase (NEB) was added and the mixture was incubated for 15 minutes at 60 °C.
[0088] Next, the ligase was heat-inactivated by incubation for 5x minutes at 94 °C and stored at -20 °C until further use.

PCR amplification:

[0089] PCR reactions mixture contained 10 µl ligation mixture, 1 µl of 50 ng/µl (FAM or ^{33}P) labelled M13 forward and reverse primer (as described in Example 4), 200 µM of each dNTP, 2.5 Units HotStarTaq Polymerase Qiagen, 5 µl 10X PCR buffer in a total volume of 50 µl.
[0090] Amplifications were carried out by thermal cycling in a Perkin Elmer 9700 thermo cycler (Perkin Elmer Cetus, Foster City, United States of America), according to the following thermal cycling profile:

Profile 1: Initial denaturation/enzyme activation 15 min at 94 °C, followed by 35 cycles of: 30 sec at 94 °C, 30 sec at 55 °C, 1 min at 72 °C, and a final extension of 2 min at 72 °C, 4 °C, forever.
Profile 2: Initial denaturation/enzyme activation 15 min at 94 °C, followed by 35 cycles of: 5 sec at 94 °C, 5 sec at 55 °C, 10 sec at 72 °C, and a final extension of, 2 min at 72 °C, 4 °C, forever.
In case a ^{33}P end-labelled M13 forward PCR primers was used, the labelling was carried out by kination as described in Vos *et al.*, 1995 (Nucleic Acids Research, vol. 23: no. 21, pp. 4407-4414, 1995 and patent EP0534858).

Example 7. Radioactive detection of 12-plex SNPWave products

[0091] Figure 10 shows an electrophoretic gel from a multiplex oligonucleotide ligation assay of the 12 Arabidopsis SNPs listed in Example 2. Following the procedures described here-in before, using DNA of the Colombia ecotype (C), Landsberg erecta ecotype (L) or a mixture of equal amount of both ecotype (C+L) as the starting material.
[0092] Figure 10 shows that the appropriate alleles of SNP's SNP SGCSNP164, SGCSNP119, SGCSNP69, SGCSNP29, SGCSNP27 and SGCSNP1 are clearly observed in the Colombia sample, , and the appropriate SNP alleles of SNP loci SGCSNP164, SGCSNP119, SGCSNP69, SGCSNP29, SGCSNP27 and SGCSNP1 are clearly observed in the Landsberg sample and that all these SNP alleles together are observed in the mixture of both samples.
[0093] This Example illustrates that at least six SNP's can be simultaneously ligated and amplified using the multiplex

ligation / amplification procedure. This example further illustrates that at least 12 SNPs can be detected in one sample. The results are represented in Table 8

Table 8

	SNP Name	Length	Allele	Result
Lan	SGCSNP164	136	C	Yes
Col	SGCSNP164	134	T	Yes
Lan	SGCSNP119	132	T	Yes
Col	SGCSNP119	130	A	Yes
Lan	SGCSNP69	128	A	Yes
Col	SGCSNP69	126	G	Yes
Lan	SGCSNP55	124	A	No
Col	SGCSNP55	122	C	Yes
Lan	SGCSNP44	120	T	No
Col	SGCSNP44	118	A	No
Lan	SGCSNP39	116	C	No
Col	SGCSNP39	114	T	Yes
Lan	SGCSNP37	112	G	Yes
Col	SGCSNP37	110	C	No
Lan	SGCSNP27	108	G	Yes
Col	SGCSNP27	106	T	Yes
Lan	SGCSNP20	104	C	Ns*
Col	SGCSNP20	102	A	Ns
Lan	SGCSNP312	104	T	Ns
Col	SGCSNP312	102	G	Ns
Lan	SGCSNP209	100	G	Yes
Col	SGCSNP209	98	C	Yes
Lan	SGCSNP1	100	A	Yes
Col	SGCSNP1	97	G	Yes

*; not scored; Col: Colombia allele, Lan: Landsberg allele

Example 8. Gel electrophoresis

[0094] Gel electrophoresis was performed as described in Vos *et al.*, Nucleic Acids research 23(21),(1995), 4407-4414. After exposure of the dried gel to phospho-imaging screens (Fuji Photo Film Co., LTD, Type BAS III) for 16 hours, an image was obtained by scanning using the Fuji scanner (Fuji Photo Film Co., LTD, Fujix BAS 2000) and stored in digital form.

Example 9. Oligonucleotide sizers for capillary electrophoresis

[0095]

sizer 94 bp:

5'fam-

ACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGA

CCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCG [SEQ ID No. 39]

sizer 95 bp:

5'fam-

ACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGA
CCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGG [SEQ ID No.40]

sizer 137 bp:

5'fam-

ACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGA
CCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGG
CATGACTATCGTCGCCGCACTTATGACTGTC [SEQ ID No.41]

Example 10. Purification and dilution of amplified connected probes

[0096] In case of detection using the MegaBACE 1000 capillary sequencing instrument, desalting and purification of the PCR reactions mixtures was carried in 96-well format, using the following procedure:

A. Preparation of the 96-well Sephadex purification plates

[0097] Dry Sephadex™ G-50 superfine (Amersham Pharmacia Biotech, Uppsala, Sweden) was loaded into the wells of a 96-well plate (MultiScreen®-HV, Millipore Corporation, Bedford, MA, USA), using the 45 microliter column loader (Millipore Corporation) as follows:

1. Sephadex G-50 superfine was added to the column loader.
2. Excess Sephadex was removed from the top of the column loader with a scraper.
3. The Multiscreen-HV plate was placed upside-down on top of the Column Loader.
4. The Multiscreen-HV plate and the Column Loader were both inverted.
5. The Sephadex G-50 was released by tapping on top or at the side of the Column Loader.

[0098] Next, the Sephadex G-50 was swollen en rinsed as follows:

6. 200 µl Milli-Q water was added per well using a multi-channel pipettor.
7. A centrifuge alignment frame was placed on top of a standard 96-well microplate, the Multiscreen-HV plate was place on top and the minicolumns were packed by centrifugation for 5 min at 900 g.
8. The 96-well plate was emptied and placed back.
9. Steps 5-7 were repeated once.
10. 200 µl Milli-Q water was added to each well to swell the Sephadex G-50 and incubated for 2-3 hours. Occasionaly, at this stage the Multiscreen-HV plates with swollen minicolumns of Sephadex G-50 superfine were tightly sealed with parafilm and stored a refrigerator at 4 °C until further use.
11. A centrifuge alignment frame was placed on top of a standard 96-well microplate, the Multiscreen-HV plate was placed on top of the assembly and the minicolumns were packed by centrifugation for 5 min at 900 g.
12. The 96-well microplate was removed.
13. The mixtures containing the amplified connected probes were carefully added to the centre of each well.
14. Using the centrifuge alignment frame, the Multiscreen-HV plate was placed on top of a new standard U-bottom microtitre plate and centrifugation was carried out for 5 min at 900 g.
15. The eluate in the standard 96-well plate (approximately 25 µl per well) contains the purified product.

B. Dilution of the purified products

[0099] Purified samples were diluted 25-75 fold in Milli-Q water before injection.

Example 11. Capillary electrophoresis on the MegaBACE

Preparation of the samples:

[0100] A 800-fold dilution of ET-900 Rox size standard (Amersham Pharmacia Biotech) was made in water. 8 µl diluted ET-900 Rox was added to 2 µl purified sample. Prior to running, the sample containing the sizing standard was heat denatured by incubation for 1 min at 94 °C and subsequently put on ice.

Detection on the MegaBACE:

[0101] MegaBACE capillaries were filled with 1X LPA matrix (Amersham Pharmacia Biotech, Piscataway, NJ, USA) according to the manufacturer's instructions. Parameters for electrokinetic injection of the samples were as follows: 45 sec at 3 kV. The run parameters were 110 min at 10 kV. Post-running, the cross-talk correction, smoothing of the peaks and cross-talk correction was carried out using Genetic Profiler software, version 1.0 build 20001017 (Molecular Dynamics, Sunnyvale, CA, USA), and electropherograms generated.

Example 12. Repeated injection on the MegaBACE.

[0102] The minimum time interval for adequate separation between two consecutively injected samples was determined by injecting the sizer sample as described in Example 8. The resulting time interval was used, with a small additional margin, when injecting the purified amplified connected probes from the oligonucleotide assay. The results are presented in Fig 11.

A. Partial electropherogram of FAM labelled detection of the Colombia sample on a capillary electrophoretic device (MegaBACE). The same multiplex mixture was injected twice. Amplified connected probes (size range 97-134 bp) and flanking sizer fragments (94, 95 and 137 bp) are all labelled with FAM
B. Partial electropherogram of FAM labelled detection of the Landsberg erecta sample on a capillary electrophoretic device (MegaBACE). The same multiplex mixture was injected twice. Amplified connected probes (size range 97-134 bp) and flanking sizer fragments (94, 95 and 137 bp) are all labelled with FAM.

Example 13. Cross-talk reduction using stuffer sequences of different lengths

[0103] In this experiment the use of different length-label combinations to avoid the negative influence of incomplete cross-talk removal on the quality of a dominantly scored (presence /absence) dataset of SNP markers is demonstrated. Stuffer lengths were chosen such that ET-ROX and JOE-labelled fragments have identical sizes, and that FAM and NED fragments have identical sizes, but differing by 1 basepair from those of ET-ROX and JOE-labelled fragments. The result is that even in case of incomplete cross-talk removal between dyes with overlapping emission spectra, the observed signal will not result in incorrect scoring because the expected sizes of the amplification products are known for every label. Hence length-label combinations define the expectance patterns for genuine signals are signals originating from incomplete cross-talk correction. The results are presented in Fig 12 and 13.

[0104] The example shows in Figure 13:

A). The effect of incomplete cross talk removal on the data quality in case of a sample that contains a ET-ROX labelled fragment of 120 basepair and a NED labelled fragment of 124 basepairs in a situation where fragments of a particular size can be observed in combination with all labels. In this case, incomplete cross-talk of ET-ROX signal into the FAM Channel at 120 bp removal leads to the incorrect scoring of a FAM fragment of 120 basepairs (in reality an ET-ROX labelled fragment of 120 basepairs).Similarly, incomplete cross-talk correction removal of NED signal into JOE at 124 bp leads to incorrect scoring of a JOE fragment of 124 basepairs (in reality a NED labelled fragment of 124 basepairs), in addition to the correct fragments.
B). The effect of the use of cross-talk-optimised length-label combinations such that ET-ROX- and FAM-labelled fragments of the same length are not avoided by choosing different stuffer lengths, because their emission spectra overlap. Similarly, same-size amplified connected probe fragments labelled with JOE and NED are avoided. In case of a hypothetical sample containing a 120 bp ET-ROX -labelled fragment and a 124 bp NED labelled fragment (identical to the that described above in A), the small but detectable signals (peaks) of FAM at 120 bp and of JOE at 124 bp that remain after incomplete (mathematical) cross-talk correction will not be scored because they are known to originate from cross talk of ET-ROX and NED signals, respectively. Hence, they have no impact on the data quality and both fragments are scored correctly.

SEQUENCE LISTING

<110> Keygene N.V.

<120> High throughput analysis and detection of multiple target sequences

<130> SNP-wave

<140> BO 44818
<141> 2001-12-13

<160> 74

<170> PatentIn Ver. 2.1

<210> 1
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 1
cgccagggtt ttcccagtca cgacttcagg actagtctat accttgag 48

<210> 2
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 2
cgccagggtt ttcccagtca cgacgacttc aggactagtc tataccttga a 51

<210> 3
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 3
ctatgtgaac caaattaaag tttactcctg tgtgaaattg ttatccgct 49

<210> 4
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 4
cgccagggtt ttcccagtca cgacctgctc tttcctcgct agcttcaga 49

<210> 5
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 5
cgccagggtt ttcccagtca cgacgactgc tctttcctcg ctagcttcag c 51

<210> 6
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 6
agattcggac ctttctctcat aatccgactt cctgtgtgaa attgttatcc gct 53

<210> 7
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 7
cgccagggtt ttcccagtca cgacgaagag gagagtggct acgaactct 49

<210> 8
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 8
cgccagggtt ttcccagtca cgacgagaag aggagagtgg ctacgaactc g 51

<210> 9
<211> 57
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 9
gcgataactg ctctgtagaa agacccgatt agctcctgtg tgaaattgtt atccgct 57

<210> 10
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 10
cgccagggtt ttcccagtca cgacaatcgg cctaagcaag cttgttttc 49

<210> 11
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 11
cgccagggtt ttcccagtca cgacgaaatc ggcctaagca agcttgtttt g 51

<210> 12
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 12
tgctattgat atctctgtgc aactcatcgg atcagcttcc tgtgtgaaat tgttatccgc 60
t 61

<210> 13
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 13
cgccagggtt ttcccagtca cgacgatcgg aaagatatcg gagctcctt 49

<210> 14
<211> 51
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 14
cgccagggtt ttcccagtca cgacgagatc ggaaagatat cggagctcct c 51

<210> 15
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 15
gtcggtgtca accgatccac ggcgcatgct agcacgtact gtcctgtgtg aaattgttat 60
ccgct 65

<210> 16
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 16
cgccagggtt ttcccagtca cgacgaactg gcatcaatca ggcctccaa 49

<210> 17
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 17
cgccagggtt ttcccagtca cgacgagaac tggcatcaat caggcctcca t 51

<210> 18
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 18
ccttaatgca agggcttatt acgtcgtacc agtcatggat ccgattcctg tgtgaaattg 60
ttatccgct 69

<210> 19
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 19
cgccagggtt ttcccagtca cgacggactc caaggtattg ttaggcgcc 49

<210> 20
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 20
cgccagggtt ttcccagtca cgacgaggac tccaaggtat tgttaggcgc a 51

<210> 21
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 21
aaccaccaag atcagtctca tcttcgatca tcgactggta ctacggactt cctgtgtgaa 60
attgttatcc gct 73

<210> 22
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 22
cgccagggtt ttcccagtca cgaccatctc ttgcgccttc tcagtgttg 49

<210> 23
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 23
cgccagggtt ttcccagtca cgacgacatc tcttgcgcct tctcagtgtt a 51

<210> 24
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 24
tgacgtccgt cgaagaatag gtaacgtacg catgctaacg ttacggacta tcgtcctgtg 60
tgaaattgtt atccgct 77

<210> 25
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 25
cgccagggtt ttcccagtca cgacagtttc aaaacccatg acgcttcta 49

<210> 26
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 26
cgccagggtt ttcccagtca cgacgaagtt tcaaaaccca tgacgcttct t 51

<210> 27
<211> 81
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 27
gtgatagctg aaaagaccca ttctccgtag catcgatatc ggtcaactgg atcagcttcc 60
tgtgtgaaat tgttatccgc t 81

<210> 28
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 28
cgccagggtt ttcccagtca cgacatactc caattgctca ggcacagtt 49

<210> 29
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 29
cgccagggtt ttcccagtca cgacgaatac tccaattgct caggcacagt c 51

<210> 30
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 30
ctccttgtcc cacgaagata gttccgtacc atgtcgacgt agatccgtat agcactgagt 60
ctcctgtgtg aaattgttat ccgct 85

<210> 31
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 31
cgccagggtt ttcccagtca cgacgtagag gctctaaaca gctgcttcc 49

<210> 32
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 32
cgccagggtt ttcccagtca cgacgagtag aggctctaaa cagctgcttc g 51

<210> 33
<211> 49

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 33
cttgtttatg ctaagggccg gctcctcctg tgtgaaattg ttatccgct 49

<210> 34
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 34
cgccagggtt ttcccagtca cgactaagtc agctcctaag cttccatcg 49

<210> 35
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 35
cgccagggtt ttcccagtca cgacgataag tcagctccta agcttccatc t 51

<210> 36
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 36
aagccacttc ctcctgctca agcgcgactt cctgtgtgaa attgttatcc gct 53

<210> 37
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 37
cgccagggtt ttcccagtca cgac 24

<210> 38
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 38
agcggataac aatttcacac agga 24

<210> 39
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 39
tgca 4

<210> 40
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 40
acgttacg 8

<210> 41
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 41
tagcgtcagc at 12

<210> 42
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 42
catggcatac gttacg 16

<210> 43
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 43
gatcgctaac gttacggcat 20

<210> 44
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 44
tcgagatcac gtcatgctga gcat 24

<210> 45
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 45
cagttcaggc attcgactag cgtatacg 28

<210> 46
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 46
gtcaatcgga ctctgagact catgcgatga ct 32

<210> 47
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 47
gatccgatcg atatcgacgt agctgcatcg taatcg 36

<210> 48
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 48
actg 4

<210> 49
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 49
gcatcagt 8

<210> 50
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 50
atcggcatta cg 12

<210> 51
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 51
tacggcatag tcacgt 16

<210> 52
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic

oligonucleotide

<400> 52
gatcgctaac gttacggcat 20

<210> 53
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 53
ctagatgctc aggctatcga catg 24

<210> 54
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 54
gtaccgatac gttagcgact tagccgta 28

<210> 55
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 55
cgtaatcgga tccgtaacgt gcatatgcca gt 32

<210> 56
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 56
gacttcgaga tctgcaacgt acgtcgtaag ctgcta 36

<210> 57
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 57
cgact 5

<210> 58
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 58
ccgattagc 9

<210> 59
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 59
catcggatca gct 13

<210> 60
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 60
catgctagca cgtactg 17

<210> 61
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 61
cgtaccagtc atggatccga t 21

<210> 62
<211> 25
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 62
cgatcatcga ctggtactac ggact 25

<210> 63
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 63
cgtacgcatg ctaacgttac ggactatcg 29

<210> 64
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 64
ccgtagcatc gatatcggtc aactggatca gct 33

<210> 65
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 65
cgtaccatgt cgacgtagat ccgtatagca ctgagtc 37

<210> 66
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 66
cgtac 5

<210> 67

<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 67
cgcattcga 9

<210> 68
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 68
catcggcatg act 13

<210> 69
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonuclectide

<400> 69
cgtcaatgca cgttacg 17

<210> 70
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 70
cgcatcgata gctctgaacg t 21

<210> 71
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 71
cgcatatcgg atcgatcgca tacgt 25

<210> 72
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 72
catcggatcc atgcgtagca tatcgacgt 29

<210> 73
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 73
ctgcaagtcc gattacgatc gacgtgctat gca 33

<210> 74
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: synthetic oligonucleotide

<400> 74
cagctcagta tcgagtcgac tacgttgcat acggatc 37

Claims

1. A method for determining the presence or absence of at least two target sequences (2) in at least two nucleic acid samples, wherein the method comprises the steps of:

(a) providing to a nucleic acid sample a pair of a first and a second oligonucleotide probe for each target sequence to be detected in the sample, whereby the first oligonucleotide probe has a section (4) at its 5'-end that is complementary to a first part (5) of a target sequence and the second oligonucleotide probe has a section (6) at its 3'-end that is complementary to a second part (7) of the target sequence, whereby the first (5) and second part (7) of the target sequence are located adjacent to each other, and whereby the first and second oligonucleotide probes (4, 6) each comprise a tag sequence (8, 9), whereby the tag sequences are essentially non-complementary to the target sequence, whereby the tag sequences may comprise a stuffer sequence (10, 11) and whereby the tag sequences comprise primer-binding sequences (12, 13);
(b) allowing the oligonucleotide probes to anneal to the adjacent parts of target sequences whereby the complementary sections (4,6) of the first and second oligonucleotide probes are adjacent;
(c) providing means (14) for connecting the first and the second oligonucleotide probes annealed adjacently

to the target sequence and allowing the complementary sections (4,6) of the adjacently annealed first and second oligonucleotide probes to be connected, to produce a connected probe (15) corresponding to a target sequence in the sample;
(d) amplifying the connected probes from a primer pair comprising a first primer (16) that is complementary to the primer-binding sequence (12) of the first oligonucleotide probe (4) and a second primer (17) that is complementary to the primer-binding sequence (13) of the second oligonucleotide probe, to produce an amplified sample (19) comprising amplified connected probes (20);
(e) repeating steps (a) to (d) to generate at least two amplified samples (19);
(f) consecutively applying at least part of the amplified samples (19) obtained in steps (d) and (e), to an application location of a channel (21) of an electrophoretic device (22), electrophoretically separating the amplified connected probes in the amplified samples (19) and detecting the separated amplified connected probes at a detection location (24) located distal from the application location of the channel; whereby the time period (23) between the consecutively applied amplified samples is such that the slowest migrating amplified connected probe (19) in an amplified sample is detected at the detection location(24), before the fastest migrating amplified connected probe of a subsequently applied amplified sample is detected at the detection location (24);
(g) determining the presence or absence of a target sequence in a sample by detecting the presence or absence of the corresponding connected probe.

2. A method according to claim 1, wherein an amplified connected probe in an amplified sample corresponding to a target sequence in a sample differs in length from an amplified connected probe in the amplified sample corresponding to a different target sequence in the sample.

3. A method according to claim 2, wherein the length difference is provided by the length of the stuffer sequence.

4. A method according to any one of claims 1 - 3, wherein the amplified connected probes corresponding to different target sequences in the sample differ in length by at least two nucleotides.

5. A method according to any one of claims 1 - 4, wherein at least one of the first and second oligonucleotide probes that are complementary to at least two different target sequences in the sample comprise a tag sequence that comprises a primer-binding site that is complementary to a single primer sequence.

6. A method according to any one of claims 1 - 5, wherein at least one of the primers complementary to the primer-binding sites of the first and second oligonucleotide probes in the sample comprises a label.

7. A method according to claim 6, wherein the label is selected from amongst the fluorescent or phosphorescent dyes selected from the group of FAM, TET, JOE, NED, HEX, (ET-)ROX, FITC, Cy2, Texas Red, TAMRA, Alexa fluor 488™, Bodipy™ FL, Rhodamine 123, R6G, Bodipy 530, Alexafluor™532 and IRDyes ™.

8. A method according to claim 1, wherein:

(a) at least two groups of pairs of first and second oligonucleotide probes are provided to a sample, whereby each group of oligonucleotide probes has tag sequences with at least one group specific primer-binding site;
(b) the connected probes of each group are amplified from primer pair wherein at least one of the two primers is complementary to the group specific primer-binding site, and whereby at least one of the primers of a group comprises a group specific label; and,
(c) in each group, an amplified connected corresponding to a target sequence in the sample, differs in length from an amplified connected probe corresponding to a different target sequence in the sample.

9. A method according to claim 8, whereby in a first part of the groups amplified connected probes are produced having an even number of nucleotides and in a second part of the groups amplified connected probes are produced having an odd number of nucleotides.

10. A methods according to claim 9, wherein the groups of connected amplified probes having an even number of nucleotides and the groups of connected amplified probes having an odd number of nucleotides are labelled with fluorescent labels which have the least overlap in their emission spectra.

11. A method according to claim 10, wherein a first and second groups of connected amplified probes having an even

number of nucleotides are produced and a third and fourth group of connected amplified probes having an odd number of nucleotides are produced and whereby the first and second group are labelled with FAM and NED, respectively, and the third and fourth group are labelled with (ET-)ROX and either JOE or HEX, respectively; or whereby the first and second group are labelled with (ET-)ROX and either JOE or HEX, respectively, and the third and fourth group are labelled with FAM and NED, respectively.

12. A method according to any one of the preceding claims, wherein the tag of the oligonucleotide probes comprise a stuffer sequence with a length of 0 to 500, preferably from 0 to 100, more preferably from 0 to 50 nucleotides.

13. A method according to any one of the preceding claims, wherein the presence or absence of at least 10, preferably at least 25, more preferably at least 50, still more preferably at least 100, most preferably at least 250 different target nucleotide sequences is determined in a nucleic acid sample.

14. A method according to any one of the preceding claims, wherein the length of the complementary section of the oligonucleotide probes is between 15 and 50 nucleotides, preferably between 18 and 40 nucleotides, more preferably between 20 and 30 nucleotides.

15. A method according to any one of the preceding claims, wherein the length of the primer-binding site is between 12 and 40 nucleotides, preferably between 15 and 30 nucleotides, more preferably between 17 and 25.

16. A method according to any one of the preceding claims, wherein the length of the tag (8,9) is between 15 and 540 nucleotides, preferably between 18 and 140 nucleotides, more preferably between 20 and 75.

17. A method according to any one of the preceding claims, wherein a length-based size standard, preferably a flanking size-ladder, is co-electrophoresed with the sample.

18. A method according to any one of the preceding claims, wherein the target nucleotide sequence contains a polymorphism, preferably a single nucleotide polymorphism.

19. A method according to any one of the preceding claims, wherein the target nucleotide sequence is a DNA molecule selected from the group consisting of: cDNA, genomic DNA, restriction fragments, adapter-ligated restriction fragments, amplified adapter-ligated restriction fragments and AFLP fragments.

20. A method according to any one of the preceding claims, further comprising a step for the removal of non-ligated probes, optionally prior to amplification, preferably by exonucleases.

21. Use of a method as defined in any of claims 1-20, for high throughput detection of a multiplicity of target nucleotide sequences.

22. Use according to claim 21, for the detection of polymorphisms, preferably single nucleotide polymorphism.

23. Use according to claim 21, for transcript profiling.

24. Use according to claim 21, for the detection of the quantitative abundance of target nucleic acid sequences.

25. Use according to claim 21, for genetic mapping, gene discovery, marker assisted selection, seed quality control, hybrid selection, QTL mapping, bulked segregant analysis, DNA fingerprinting and for disclosing information relating to traits, disease resistance, yield, hybrid vigour, and/or gene function.

26. A oligonucleotide acid probe for use in a method as defined in claims 1-20.

27. A set of two or more oligonucleotide probes, for use in a method as defined in claims 1-20.

28. Use of a set of two or more oligonucleotide probes as defined in claim 27, wherein the set comprises a probe for each allele of a single nucleotide polymorphism.

29. A set of primers for use in a method according to any one of claims 1-20, the set comprising a first primer and one or more second primers, wherein each second primer contains a label.

30. A kit comprising oligonucleotide probes suitable for use in a method as defined in claims 1-20.

31. A kit comprising primers for use in a method as defined in claims 1-20.

32. A kit comprising primers and oligonucleotide probes for use in a method as defined in claims 1-20.

Fig 1

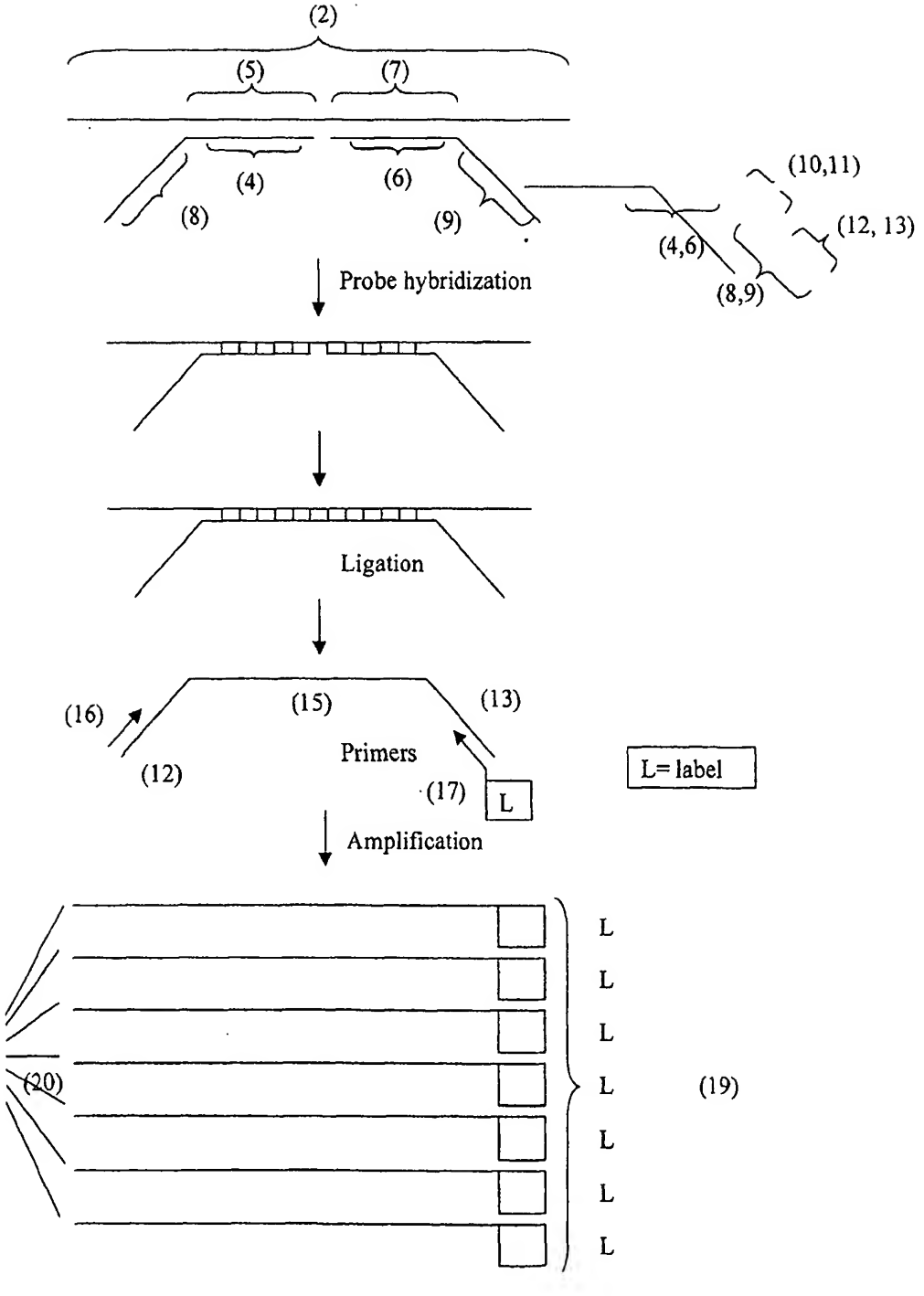

(2)
(5)
(7)
(4)
(6)
(8)
(9)
(10,11)
(12, 13)
(4,6)
(8,9)
Probe hybridization
Ligation
(16)
(15)
(13)
Primers
(12)
(17)
L
L= label
Amplification
L
L
L
L
L
L
L
(20)
(19)

Fig 2

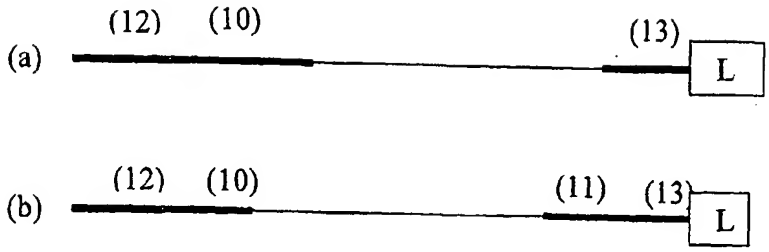
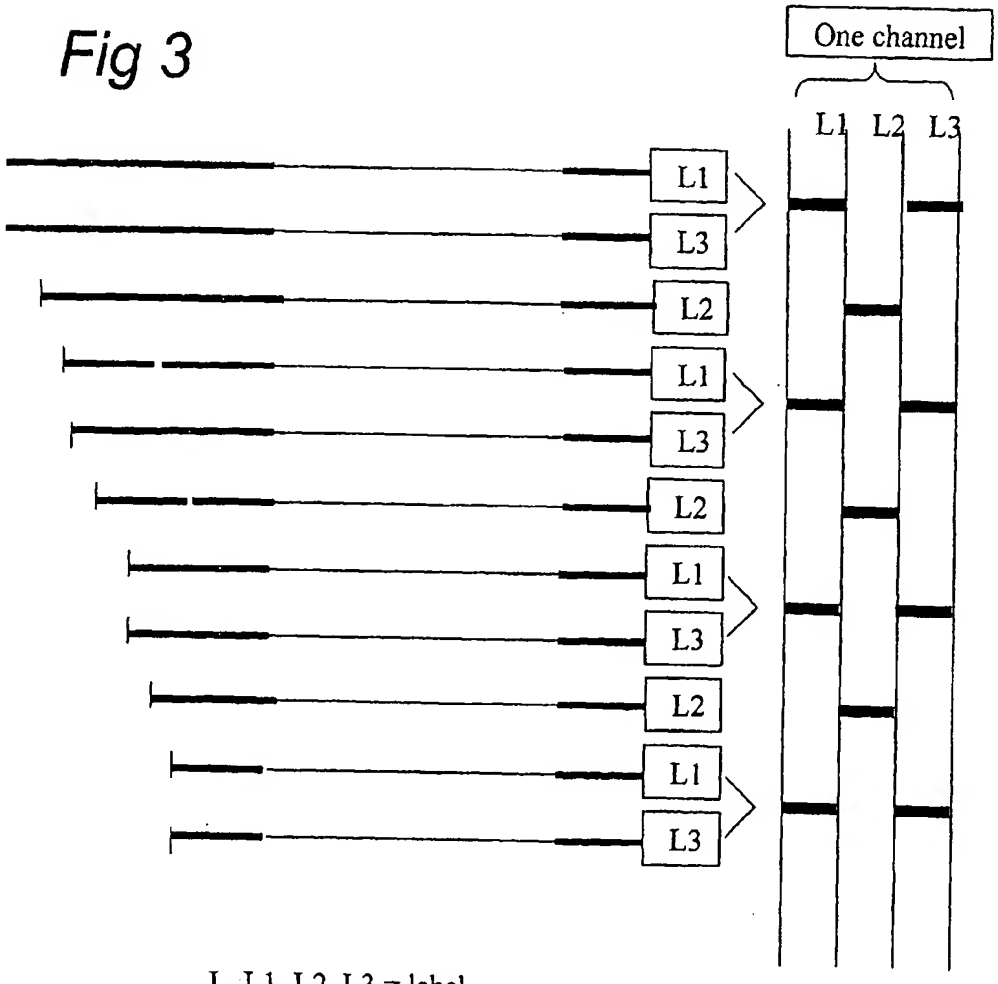


Fig 3



L, L1, L2, L3 = label

Fig 4

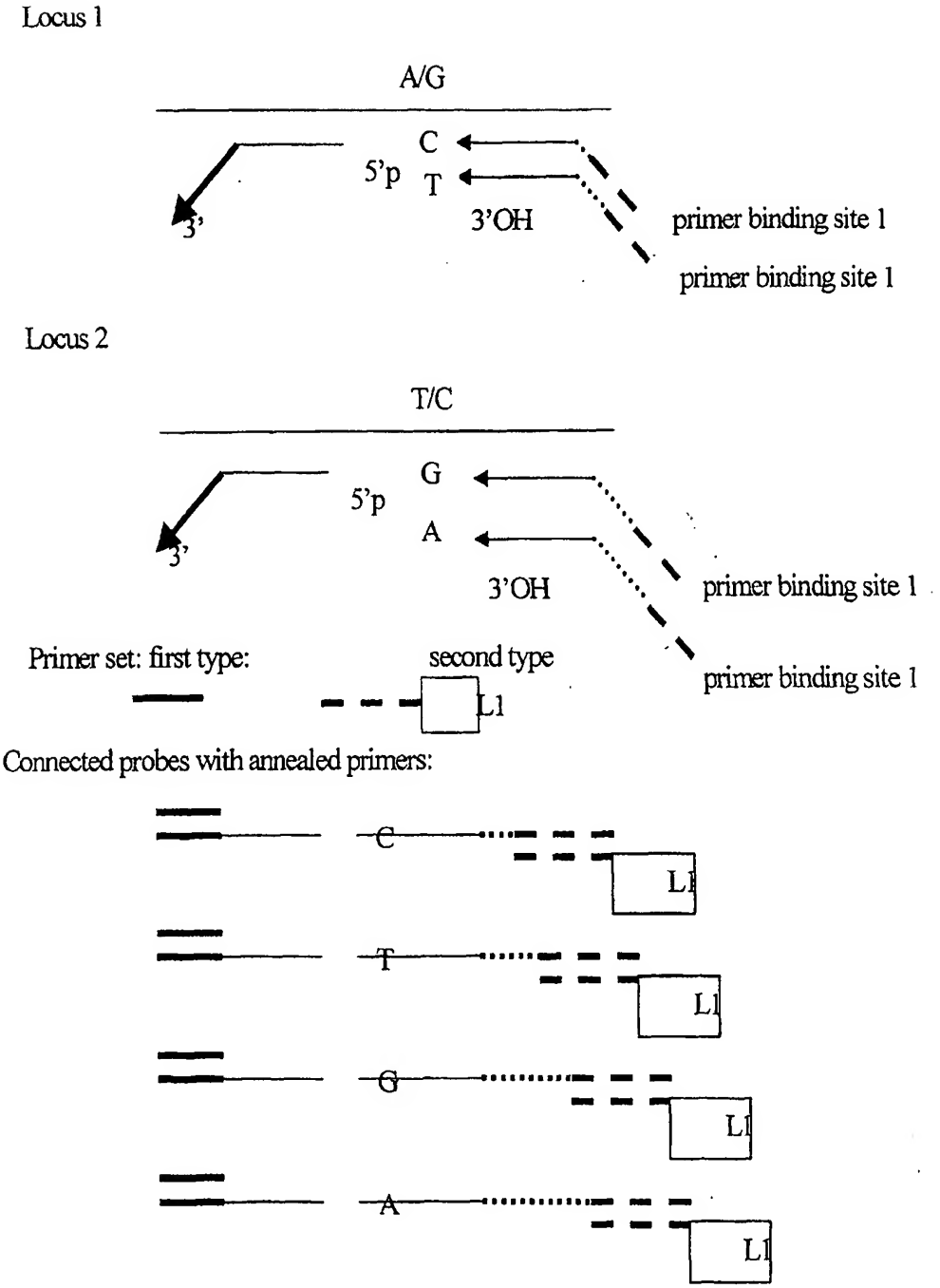
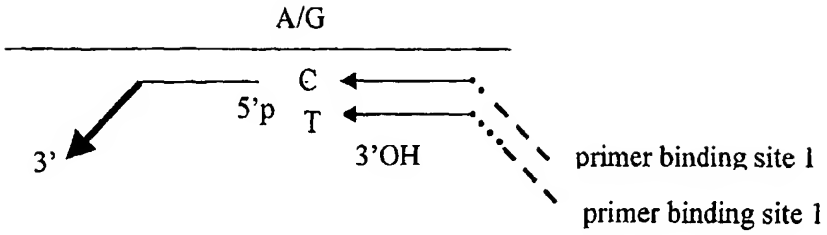

Locus 1
A/G
C
5'p
T
3'
3'OH
primer binding site 1
primer binding site 1
Locus 2
T/C
G
5'p
A
3'
3'OH
primer binding site 1
primer binding site 1
Primer set: first type:
second type
L1
Connected probes with annealed primers:
C
L1
T
L1
G
L1
A
L1

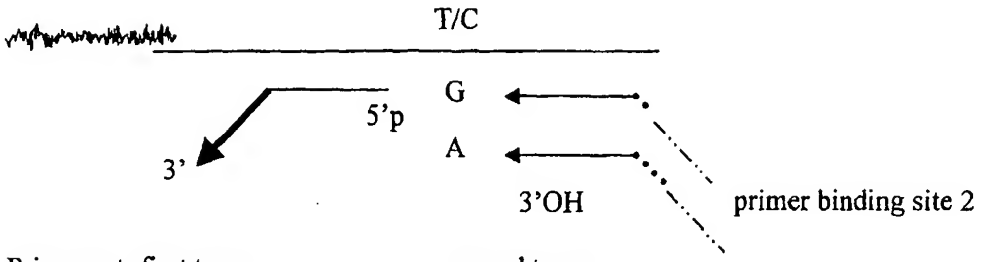
Fig 5

a

Locus 3



Locus 4



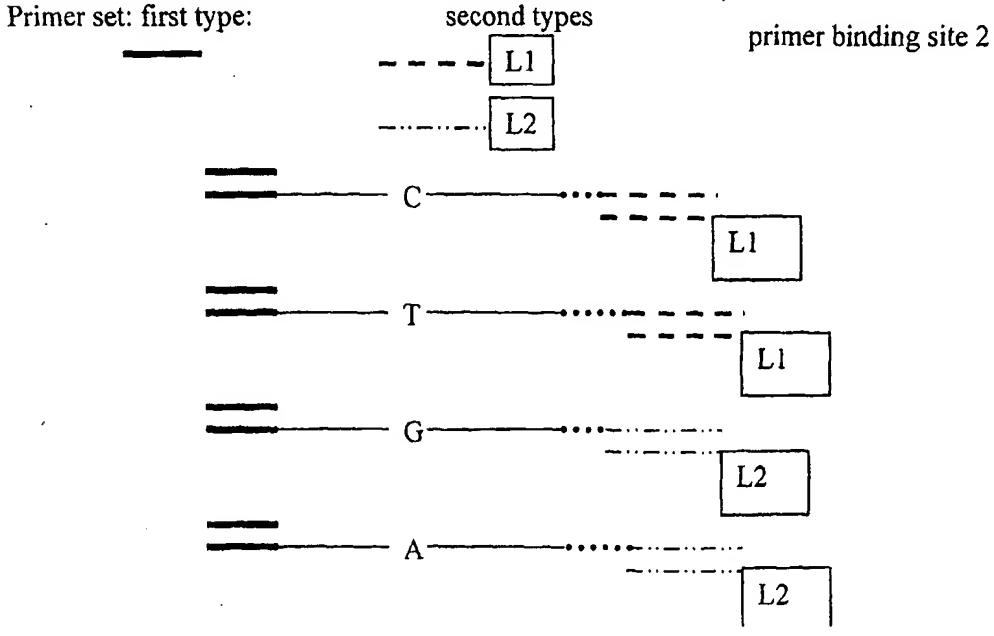


Fig 6

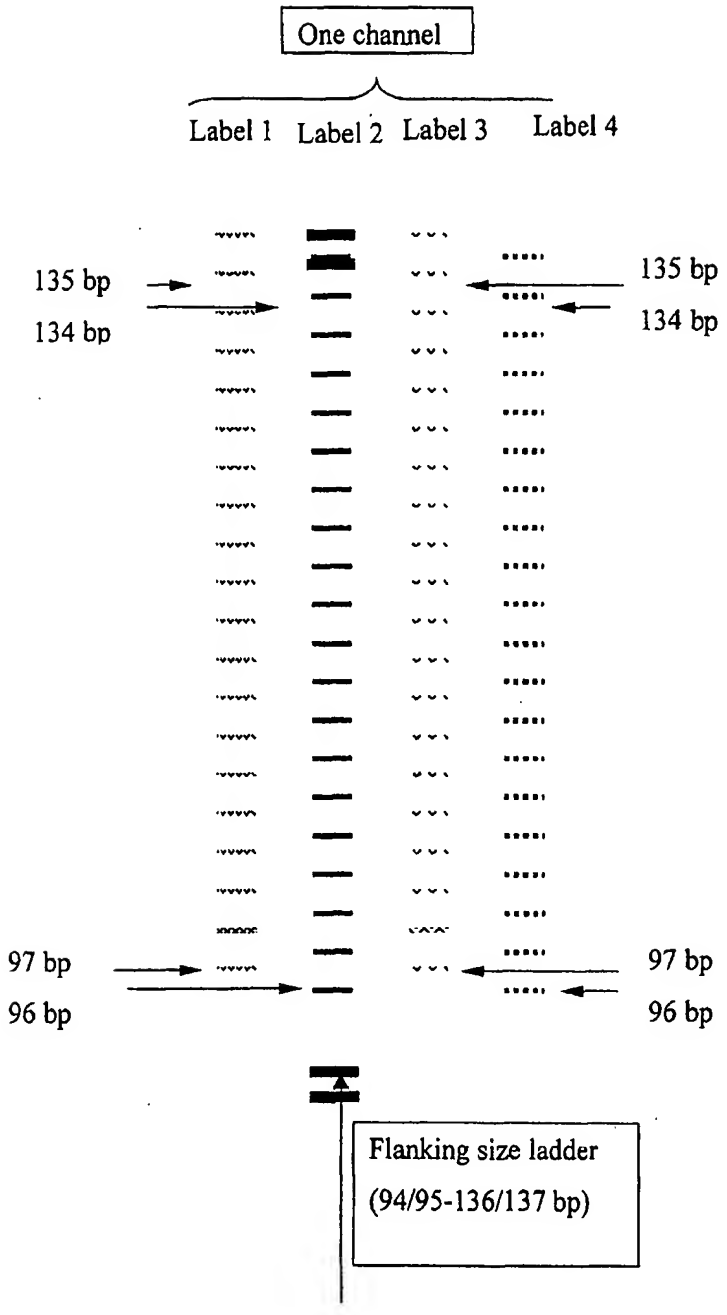

One channel
Label 1
Label 2
Label 3
Label 4
135 bp
134 bp
135 bp
134 bp
97 bp
96 bp
97 bp
96 bp
Flanking size ladder
(94/95-136/137 bp)

Fig 7

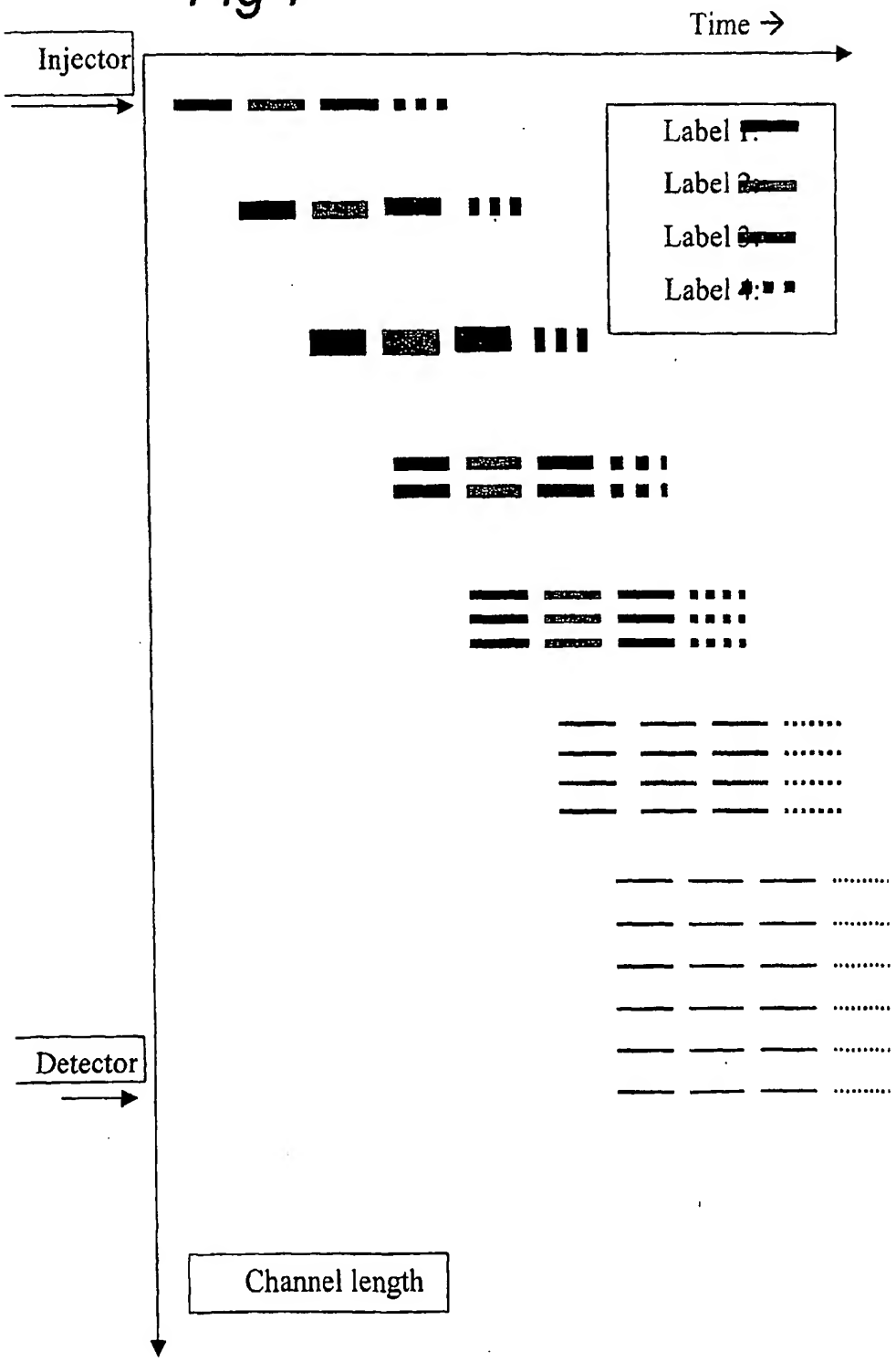

Time →
Injector
Label 1
Label 2
Label 3
Label 4
Detector
Channel length

Fig 8

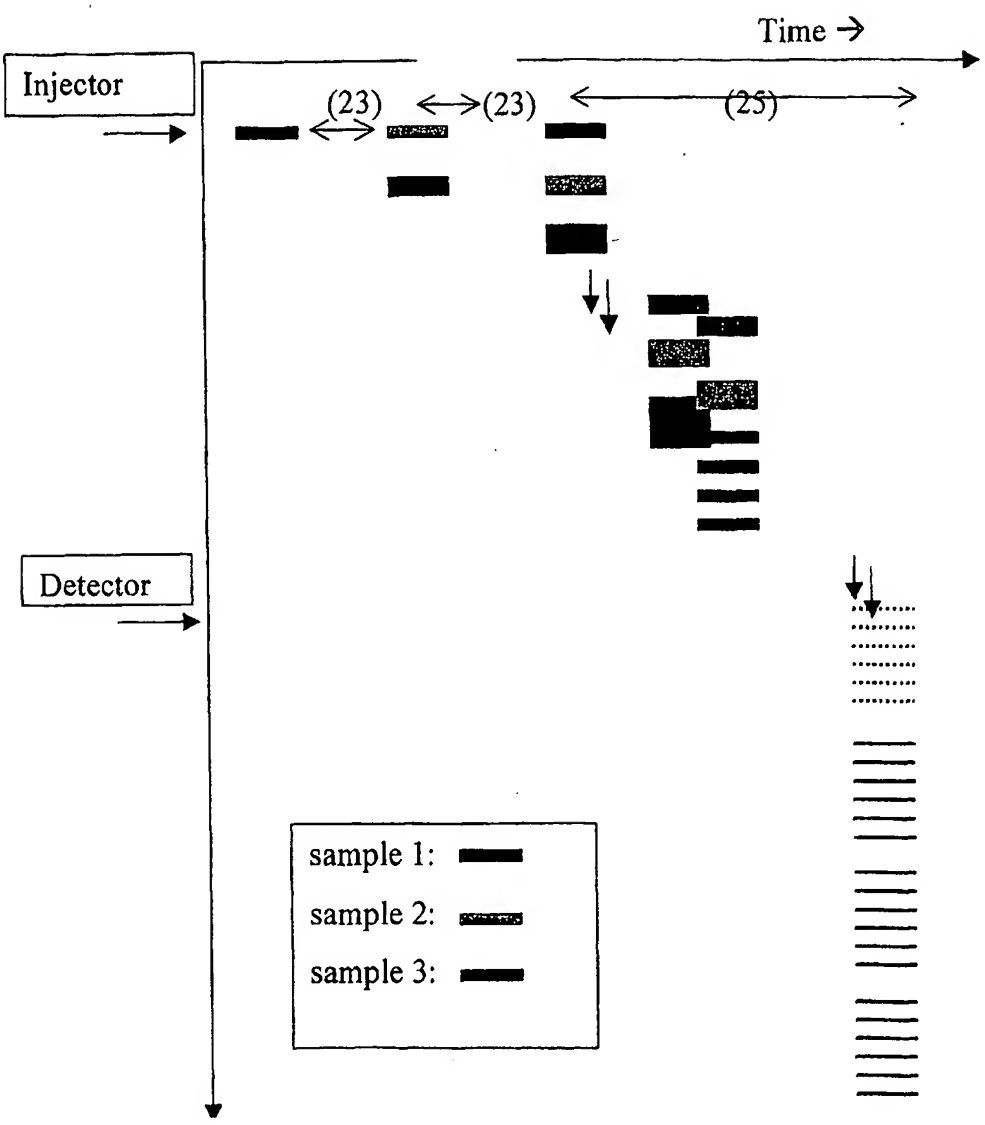


Fig 9

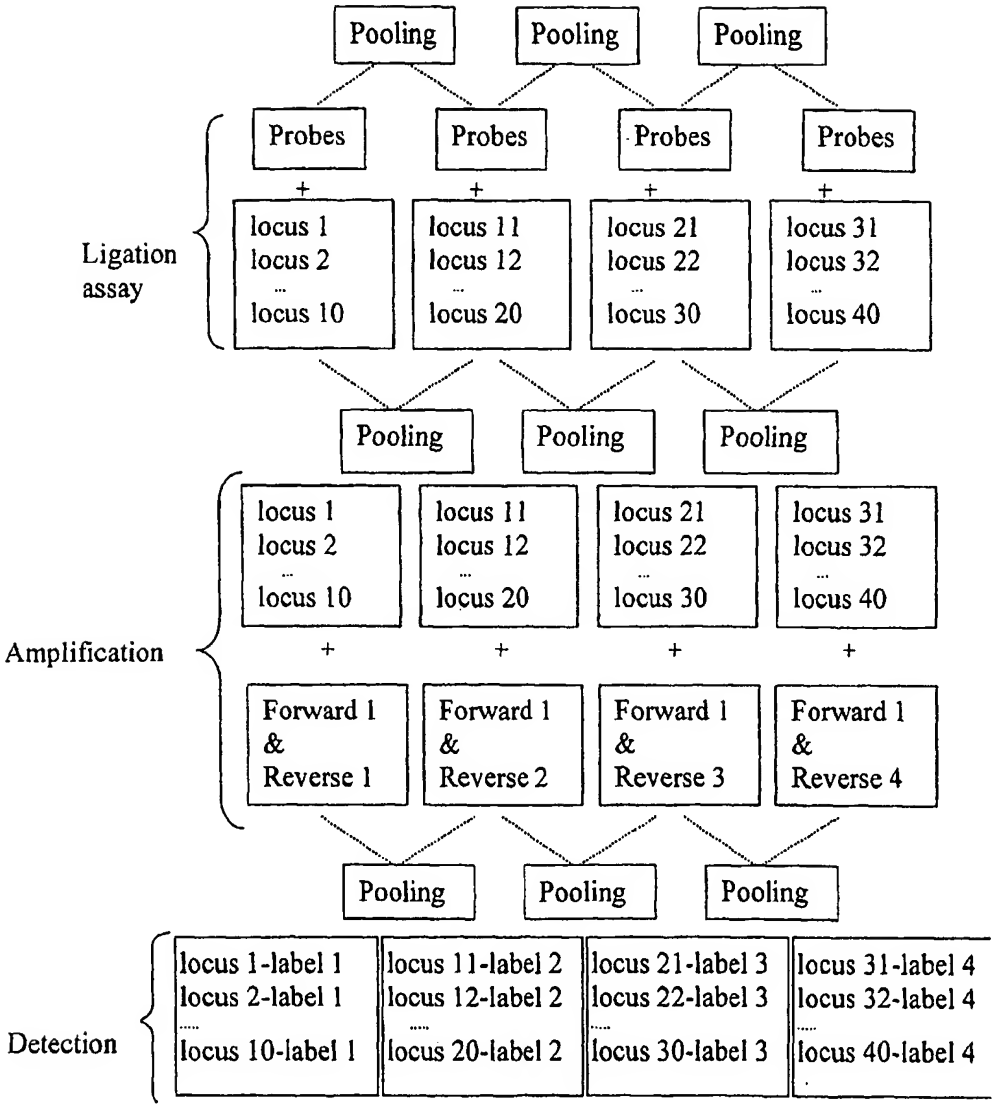

Pooling
Pooling
Pooling
Probes
Probes
Probes
Probes
+
+
+
+
Ligation assay
locus 1
locus 2
...
locus 10
locus 11
locus 12
...
locus 20
locus 21
locus 22
...
locus 30
locus 31
locus 32
...
locus 40
Pooling
Pooling
Pooling
Amplification
locus 1
locus 2
...
locus 10
locus 11
locus 12
...
locus 20
locus 21
locus 22
...
locus 30
locus 31
locus 32
...
locus 40
+
+
+
+
Forward 1 & Reverse 1
Forward 1 & Reverse 2
Forward 1 & Reverse 3
Forward 1 & Reverse 4
Pooling
Pooling
Pooling
Detection
locus 1-label 1
locus 2-label 1
....
locus 10-label 1
locus 11-label 2
locus 12-label 2
....
locus 20-label 2
locus 21-label 3
locus 22-label 3
....
locus 30-label 3
locus 31-label 4
locus 32-label 4
....
locus 40-label 4

Fig 10

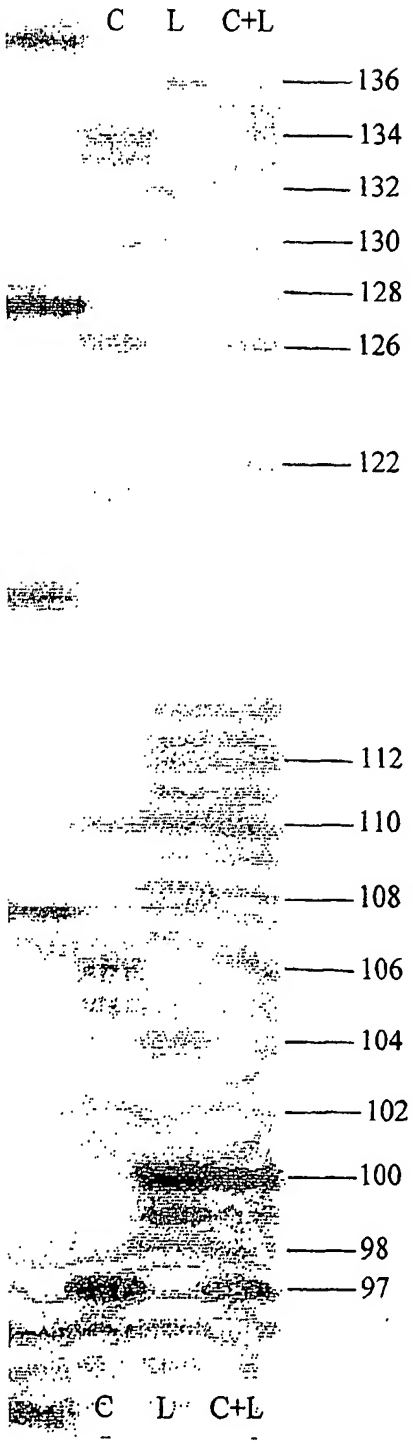

C
L
C+L
136
134
132
130
128
126
122
112
110
108
106
104
102
100
98
97
C
L
C+L

Fig 11

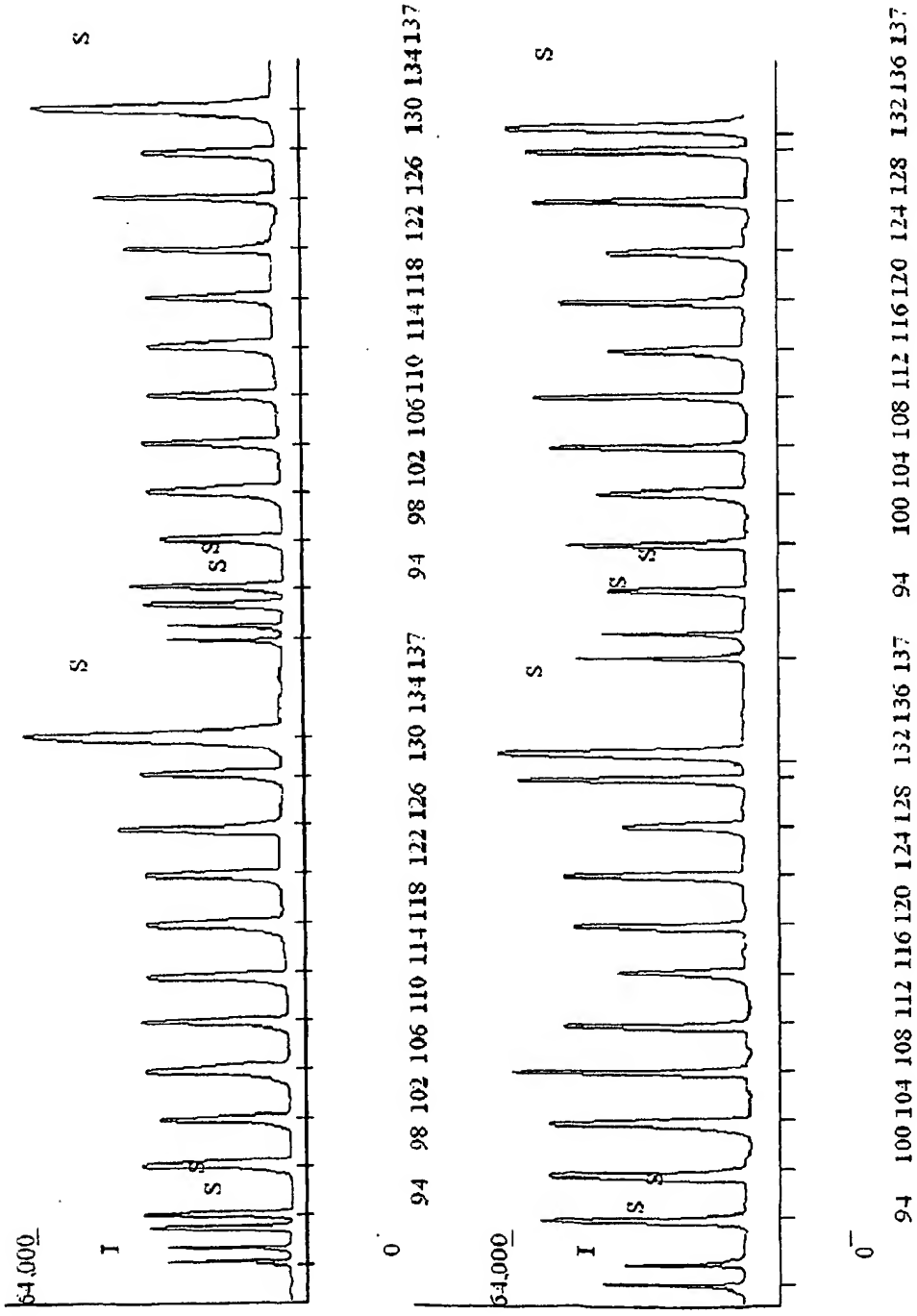

64,000
I
S
S
S
S
S
0
94 98 102 106 110 114 118 122 126 130 134 137
64,000
I
S
S
S
S
S
0
94 100 104 108 112 116 120 124 128 132 136 137

Fig 12

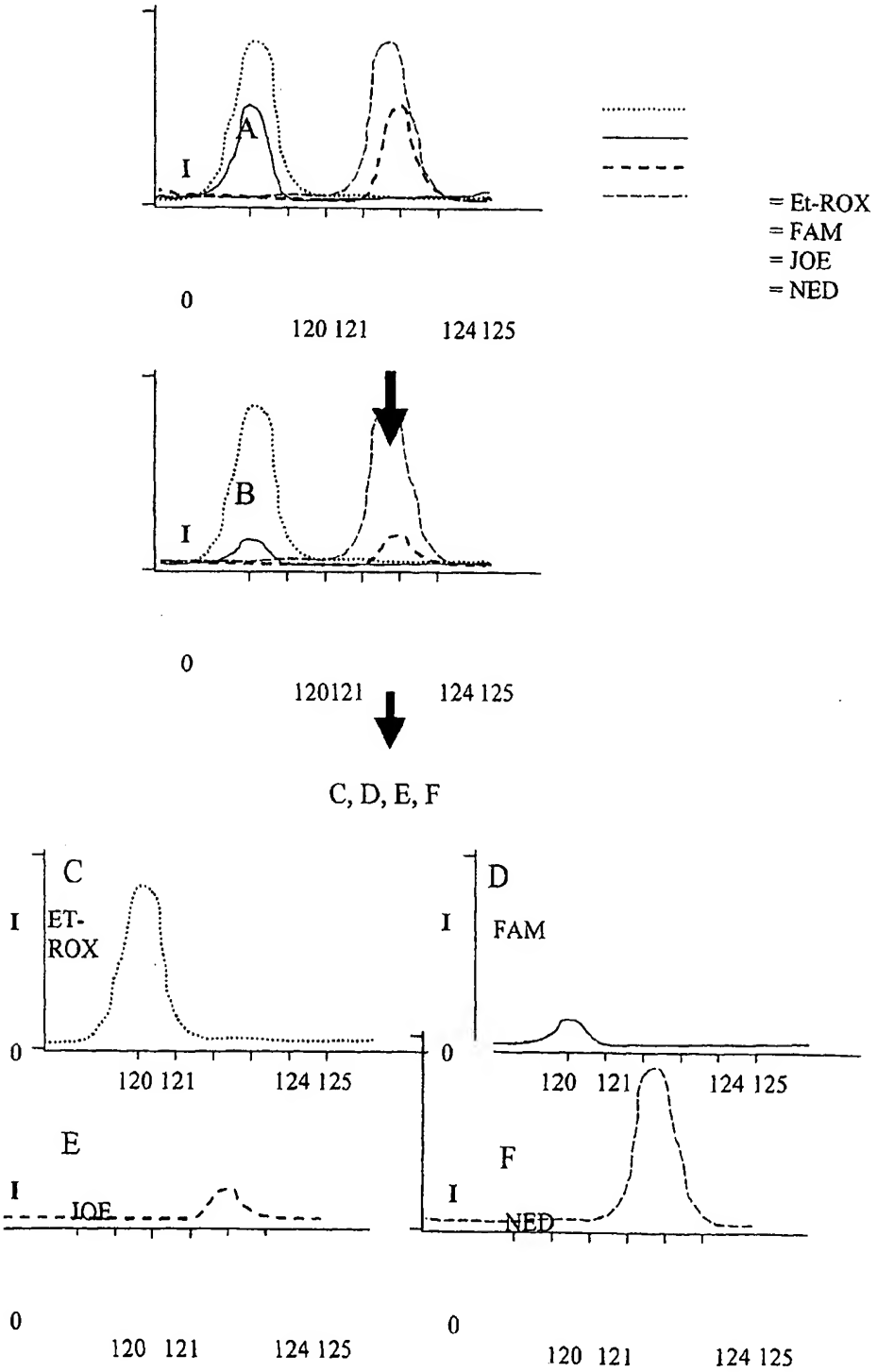

A
I
= Et-ROX
= FAM
= JOE
= NED
0
120 121
124 125
B
I
0
120121
124 125
C, D, E, F
C
I
ET-
ROX
0
120 121
124 125
D
I
FAM
0
120 121
124 125
E
I
JOE
0
120 121
124 125
F
I
NED
0
120 121
124 125

Fig 13

A	Observed data		Scored data		Expected data	
bp	**120**	**124**	**120**	**124**	**120**	**124**
ET-ROX	+	-	+	-	+	-
FAM	+	-	⊕	-	-	-
JOE	-	+	-	⊖	-	-
NED	-	+	-	+	-	+

B	Observed data				Scored data				Expected data			
bp	**120**	**121**	**124**	**125**	**120**	**121**	**124**	**125**	**120**	**121**	**124**	**125**
ET-ROX	+	-	-	-	+		-		+		-	
FAM	+	-	-	-		-		-		-		-
JOE	-	-	+	-		-		-		-		-
NED	-	-	+	-	-		+		-		+	

Legend	
+	Present
-	Absent
⊖	False (Present/Absent)
	Signal ignored

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 20 4912

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X Y	EP 1 130 113 A (SCHOUTEN JOHANNES PETRUS) 5 September 2001 (2001-09-05) * page 7, line 25 - line 51 * * page 14, line 18 - line 25; figure 2; example 1 * * page 14, line 46 - page 15, line 26; figure 1 *	26-32 1-25	C12Q1/68
D,X Y	WO 97 45559 A (BELGRADER PHILLIP ;CORNELL RES FOUNDATION INC (US)) 4 December 1997 (1997-12-04) * page 45, line 22 - page 46, line 21; figures 12,13 *	26-32 1-25	
X Y	WO 96 15271 A (ABBOTT LAB) 23 May 1996 (1996-05-23) * page 11, line 26 - page 17, line 34; example 2 *	26-32 1-25	
D,Y	US 6 156 178 A (PEPONNET CHRISTINE ET AL) 5 December 2000 (2000-12-05) * column 5, line 18 - column 6, line 2; figure 2 * * column 1, line 1 - column 5, line 16 *	1-25	TECHNICAL FIELDS SEARCHED (Int.Cl.7) C12Q
Y	GROSSMAN ET AL: "HIGH-DENSITY MULTIPLEX DETECTION OF NUCLEIC ACID SEQUENCES: OLIGONUCLEOTIDE LIGATION ASSAY AND SEQUENCE-CODED SEPARATION" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 22, no. 21, 1994, pages 4527-4534, XP002144258 ISSN: 0305-1048 * Abstract, Introduction, Materials and Methods *	1-25	

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	2 July 2002	Aguilera, M

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 01 20 4912

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
A	BARANY F: "GENETIC DISEASE DETECTION AND DNA AMPLIFICATION USING CLONED THERMOSTABLE LIGASE" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 88, no. 1, 1991, pages 189-193, XP000368693 ISSN: 0027-8424 * whole document *	1-32	
A	LANDEGREN U ET AL: "A LIGASE-MEDIATED GENE DETECTION TECHNIQUE" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, US, vol. 241, no. 4869, 26 August 1988 (1988-08-26), pages 1077-1080, XP000676556 ISSN: 0036-8075 * whole document *	1-32	
A	SHI MICHAEL M: "Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY. WINSTON, US, vol. 47, no. 2, February 2000 (2000-02), pages 164-172, XP002197957 ISSN: 0009-9147 * page 167; figure 3 *	1-32	TECHNICAL FIELDS SEARCHED (Int.Cl.7)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	2 July 2002	Aguilera, M

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 01 20 4912

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

02-07-2002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1130113 A	05-09-2001	EP 1130113 A1	05-09-2001
		AU 4643901 A	27-08-2001
		WO 0161033 A2	23-08-2001
WO 9745559 A	04-12-1997	AU 730633 B2	08-03-2001
		AU 3216097 A	05-01-1998
		EP 0912761 A1	06-05-1999
		JP 2000511060 T	29-08-2000
		WO 9745559 A1	04-12-1997
		US 6268148 B1	31-07-2001
		US 6027889 A	22-02-2000
WO 9615271 A	23-05-1996	WO 9615271 A1	23-05-1996
US 6156178 A	05-12-2000	AU 5928900 A	30-01-2001
		EP 1194770 A1	10-04-2002
		WO 0104618 A1	18-01-2001
		US 2001053554 A1	20-12-2001

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82